GLOBAL WELL-POSEDNESS FOR THE YANG-MILLS EQUATION IN $4+1$ DIMENSIONS. SMALL ENERGY.

JOACHIM KRIEGER AND DANIEL TATARU

Abstract. We consider the hyperbolic Yang-Mills equation on the Minkowski space $\mathbb{R}^{4+1}$. Our main result asserts that this problem is globally well-posed for all initial data whose energy is sufficiently small. This solves a longstanding open problem.

1. Introduction

Let $\mathbf{G}$ be a semisimple Lie group and $\mathfrak{g}$ its associated Lie algebra. We denote by $ad(X)Y = [X,Y]$ the Lie bracket on $\mathfrak{g}$ and by $\langle X,Y\rangle = tr(ad(X)ad(Y))$ its associated nondegenerate Killing form. The action of $\mathbf{G}$ on $\mathfrak{g}$ by conjugation is denoted by $Ad(O)X = OXO^{-1}$. We recall that the Killing form is invariant, in the sense that

$$\langle [X,Y], Z\rangle = \langle X, [Y,Z]\rangle, \qquad X,Y,Z\in\mathfrak{g},$$

or equivalently

$$\langle X,Y\rangle = \langle Ad(O)X, Ad(O)Y\rangle, \qquad X,Y\in\mathfrak{g}, \quad O\in\mathbf{G}.$$

Let $\mathbb{R}^{4+1}$ be the five dimensional Minkowski space equipped with the standard Lorentzian metric $m = \mathrm{diag}(-1,1,1,1,1)$. Denote by $A_\alpha : \mathbb{R}^{4+1}\to\mathfrak{g}$, $\alpha = 1\dots,4$, a connection form taking values in the Lie algebra $\mathfrak{g}$, and by D_α the associated covariant differentiation,

$$D_\alpha B := \partial_\alpha B + [A_\alpha, B],$$

acting on $\mathfrak{g}$ valued functions B. Introducing the curvature tensor

$$F_{\alpha\beta} := \partial_\alpha A_\beta - \partial_\beta A_\alpha + [A_\alpha, A_\beta],$$

the *Yang-Mills* equations are the Euler-Lagrange equations associated with the formal Lagrangian action functional

$$\mathcal{L}(A_\alpha,\phi) := \frac{1}{2}\int_{\mathbb{R}^{4+1}} \langle F_{\alpha\beta}, F^{\alpha\beta}\rangle\, dxdt.$$

Here we are using the standard convention for raising indices. Thus, the Yang-Mills equations take the form

$$D^\alpha F_{\alpha\beta} = 0. \tag{1.1}$$

There is a natural energy-momentum tensor associated to the Yang-Mills equations, namely

$$T_{\alpha\beta} = \frac{1}{2}m^{\gamma\delta}\langle F_{\alpha\gamma}, F_{\delta\beta}\rangle - \frac{1}{4}m_{\alpha\beta}\langle F_{\gamma\delta}, F^{\gamma\delta}\rangle.$$

The first author was partially supported by the Swiss National Science Foundation. The second author was supported in part by the NSF grant DMS-1266182 as well as by a Simons Investigator grant from the Simons Foundation.

If A solves the Yang-Mills equations (1.1) then $T_{\alpha\beta}$ is divergence free,

$$\partial^\alpha T_{\alpha\beta} = 0. \tag{1.2}$$

Integrating this for $\beta = 0$ yields a conserved energy

$$E(A) = \int_{\mathbb{R}^4} T_{00}\, dx \approx \|F\|_{L^2}^2. \tag{1.3}$$

The case $\beta \neq 0$ yields further conservation laws, i.e. the momentum, which play no role in the present article.

The Yang-Mills equations also have a scale invariance property,

$$A(t,x) \to \lambda A(\lambda t, \lambda x).$$

The energy functional E is invariant with respect to scaling precisely in dimension $4+1$. For this reason we call the $4+1$ problem energy critical; this is one of the motivations for our interest in this problem.

In order to study the Yang-Mills equations as well-defined evolutions in time we first need to address its gauge invariance. Precisely, the equations (1.1) are invariant under the gauge transformations

$$A_\alpha \longrightarrow O A_\alpha O^{-1} - \partial_\alpha O O^{-1},$$

with O elements of the corresponding group G. In order to uniquely determine the solutions to the Yang-Mills equations we need to add an additional set of constraint equations which uniquely determine the gauge. This procedure is known as *gauge fixing*.

To motivate our choice we introduce the covariant wave operator

$$\Box_A := D^\alpha D_\alpha.$$

Then we can write the Yang-Mills system in the following form

$$\Box_A A_\beta = D^\alpha \partial_\beta A_\alpha = \partial_\beta \partial^\alpha A_\alpha + [A^\alpha, \partial_\beta A_\alpha]. \tag{1.4}$$

Expanded out, the equations take the form

$$\Box A_\beta - \partial_\beta \partial^\alpha A_\alpha + \partial^\alpha [A_\alpha, A_\beta] + [A^\alpha, \partial_\alpha A_\beta - \partial_\beta A_\alpha + [A_\alpha, A_\beta]] = 0,$$

or

$$\Box A_\beta + 2[A_\alpha, \partial^\alpha A_\beta] = \partial_\beta \partial^\alpha A_\alpha - [\partial^\alpha A_\alpha, A_\beta] + [A^\alpha, \partial_\beta A_\alpha] - [A^\alpha, [A_\alpha, A_\beta]].$$

A natural condition which insures that the above system is strictly hyperbolic is the Lorenz gauge, $\partial^\alpha A_\alpha = 0$. Unfortunately there are multiple technical difficulties if one tries to implement such a gauge in the low regularity setting, see e.g. [26]. For this reason we will instead impose the *Coulomb Gauge condition* which requires

$$\sum_{j=1}^4 \partial_j A_j = 0. \tag{1.5}$$

We remark that a somewhat similar gauge is the *temporal gauge*, namely $A_0 = 0$. Another choice which is likely better but more involved technically is the caloric gauge, see e.g. [19].

Returning to the Coulomb gauge, we can use it to view the equations as a nonlocal hyperbolic system for the spatial components A_j; precisely, they solve the system

$$\Box_A A_j = \partial_j \partial_t A_0 + [A^\alpha, \partial_j A_\alpha].$$

In order to eliminate the first term on the right and also to restrict the evolution to divergence free fields A_j we apply the Leray projection $\mathbf{P}$, and rewrite the equation in the form

$$\Box A_j = \mathbf{P}\left([A^\alpha, \partial_j A_\alpha] - 2[A^\alpha, \partial_\alpha A_j] - [\partial_0 A_0, A_j] - [A^\alpha, [A_\alpha, A_j]]\right). \tag{1.6}$$

The nonlocality is due to the A_0 component, which solves an elliptic equation at fixed time, namely

$$\Delta_A A_0 = [A_j, \partial_0 A_j]. \tag{1.7}$$

The time derivative of A_0 also appears in the A_j system, so it is useful to derive an equation for it as well. This has the form

$$\Delta \partial_0 A_0 = \partial_0 \partial_j [A_0, A_j]. \tag{1.8}$$

To summarize, in the Coulomb gauge, the Yang-Mills system can be cast in the following expanded out form:

$$\begin{aligned} \Box A_i + 2[A_\alpha, \partial^\alpha A_i] &= \partial_i \partial_t A_0 + [\partial_0 A_0, A_i] + [A^\alpha, \partial_i A_\alpha] - [A^\alpha, [A_\alpha, A_i]], \\ \Delta A_0 + 2[A_i, \partial_i A_0] &= [A_i, \partial_0 A_i] - [A_i, [A_i, A_0]]. \end{aligned}$$

We will consider the solvability question for the system (1.6) in the class of divergence free vector fields, with initial data at time $t = 0$,

$$(A_j(0), \partial_0 A_j(0)) = (A_{0j}, A_{1j}) \in \mathcal{H} := \dot{H}^1(\mathbb{R}^4) \times L^2(\mathbb{R}^4). \tag{1.9}$$

We will also consider higher regularity properties of the solutions, using the spaces

$$\mathcal{H}^N := (\dot{H}^N(\mathbb{R}^4) \cap \dot{H}^1(\mathbb{R}^4)) \times H^{N-1}(\mathbb{R}^4), \qquad N \geq 1$$

Here the dependent variables A_0, $\partial_0 A_0$ are determined by the linear equations (1.7), (1.8). We remark that the solvability for these equations in various spaces, including $\dot{H}^1 \times L^2$ at fixed time, is considered in Section 2.

In order to study the dependence of the solutions on the initial data we will also need the linearized Yang-Mills equation,

$$\begin{aligned} \Box B_j = \mathbf{P}\big(&[A^\alpha, \partial_j B_\alpha] - 2[A^\alpha, \partial_\alpha B_j] - 2[B^\alpha, \partial_\alpha A_j] - [\partial_0 A_0, B_j] - [\partial_0 B_0, A_j] \\ &-2[B^\alpha, [A_\alpha, A_j]] - [A^\alpha, [A_\alpha, B_j]]\big) \end{aligned} \tag{1.10}$$

with appropriate linear elliptic equations for B_0, $\partial_0 B_0$,

$$\Delta_A B_0 = [B_j, \partial_0 A_j] + [A_j, \partial_0 B_j] + 2[B_j, \partial_j A_0] + 2[B_j, [B_j, A_0]], \tag{1.11}$$

$$\Delta \partial_0 B_0 = \partial_0 \partial_j([B_0, A_j] + [A_0, B_j]). \tag{1.12}$$

For the linearized equation we will go below scaling in regularity, and use the spaces

$$\dot{\mathcal{H}}^s = \dot{H}^s(\mathbb{R}^4) \times \dot{H}^{s-1}(\mathbb{R}^4),$$

with $s < 1$ but close to 1. Now we can state our main result:

Theorem 1. *The Yang-Mills system in Coulomb gauge* (1.6)-(1.7)-(1.8) *is globally well-posed in $\mathcal{H}$ for initial data which is small in $\mathcal{H}$, in the following sense:*

(i) (Regular data) If in addition the data (A_{0j}, A_{1j}) is more regular, $(A_{0j}, A_{1j}) \in \mathcal{H}^N$, then there exists a unique global regular solution $(A_j, \partial_0 A_j) \in C(\mathbb{R}, \mathcal{H}^N)$, which has a Lipschitz dependence on the initial data locally in time in the $\mathcal{H}^N$ topology.

(ii) (Rough data) The flow map admits an extension

$$\mathcal{H} \ni (A_{j0}, A_{j1}) \to (A_j, \partial_t A_j) \in C(\mathbb{R}, \mathcal{H})$$

within the class of initial data which is small in $\mathcal{H}$, and which is continuous in the $\mathcal{H} \cap \dot{\mathcal{H}}^s$ topology for $s < 1$ and close to 1.

(iii) (Weak Lipschitz dependence) The flow map is globally Lipschitz in the $\dot{\mathcal{H}}^s$ topology for $s < 1$, close to 1.

To clarify, in part (ii) the $\mathcal{H} \cap \dot{\mathcal{H}}^s$ norm is applied to differences of solutions. In particular, we remark that $\mathcal{H}^N$ is dense in $\mathcal{H}$ in this topology, so this extension yields solutions for all small data in $\mathcal{H}$. The $\dot{\mathcal{H}}^s$ norm plays an essential role here, as this is the norm where we have Lipschitz dependence of the solutions on the initial data. If we limit ourselves to just the $\mathcal{H}$ topology, then the best we can prove is a local in time continuous dependence on data; thus, the scattering information is lost.

We remark that in effect the proof of the theorem provides a stronger statement, where the regularity of the solutions is described in terms of function spaces S^1, S^N which incorporate both Strichartz norms, $X^{s,b}$ norms and null frame spaces. For convenience, the stronger result is stated later in Theorem 2.

Implicit in Theorem 2 is also a scattering result; however, this is not so easy to state as it is a modified rather than linear scattering. In a weaker sense, one can think of scattering as simply the fact that the S^1 norm is finite.

1.1. **Brief historical remarks.** The Yang-Mills equation belongs to the larger class of geometric nonlinear wave equations, which includes other problems such as Wave-Maps and the (mass-less) Maxwell-Klein-Gordon system. These problems have a number of shared features, including the gauge structure, and the null condition. Also, in all these problems the nonlinearity is nonperturbative at critical scaling, though only mildly so, more precisely in a way which can be addressed via renormalization. For these reasons, the understanding of these problems has evolved in a related fashion, and, as we describe below, our work on Yang-Mills was strongly influenced by prior developments for both Wave-Maps and Maxwell-Klein-Gordon.

For the Yang-Mills equation, a first global regularity result on a Minkowski background in the physical dimension $n = 3$ was first established for large data in classical work by Eardley-Moncrief, [6], [7], after earlier work by Choquet-Bruhat and Christodoulou had proved a small data global existence result in [4]. The physical $n = 3$ case is energy subcritical, which makes this problem easier from the point of view of global existence than the critical case $n = 4$, but harder from the point of view of understanding scattering.

The Eardley-Moncrief result was revisited and significantly strengthened by Klainerman-Machedon [10]. In fact, these authors showed local (and thence global) well-posedness in H^1. This work proved important for future developments on account of the fact that it identified the null-structure and its use via bilinear null-form estimates, which is also of paramount importance in this work. The energy critical case $n = 4$ of the Yang-Mills system was first attacked in Klainerman-Tataru [11]; more precisely, a model system with similar null-structures was considered there, and almost optimal local well-posedness (in light of the scaling of the system) was shown. Somewhat later, Machedon-Sterbenz [17] revisited the closely related subcritical Maxwell-Klein Gordon system in $3+1$ dimensions, and exploiting a deep trilinear null-structure in the system, managed to push local well-posedness all the

way to an almost optimal $H^{\frac{1}{2}+\epsilon}$-result (optimal in light of scaling). The new null-structure used there will also be of fundamental importance for our work.

Further work on the Maxwell-Klein-Gordon and Yang-Mills equation followed in the wake of important progress on the Wave Maps equation by the second author in [36, 37] as well as by Tao in [30]. These works introduced the functional framework that will be crucial for the present paper. In [23], Rodnianski and Tao established an optimal small data global existence result at the scaling invariant level for high-dimensional Maxwell-Klein-Gordon in the Coulomb Gauge. The important innovation there was the use of an approximate parametrix for a magnetic potential wave equation to deal with certain bad interaction terms which could not be handled perturbatively. By refining this and working with more sophisticated Banach spaces coming from the theory of Wave Maps, the authors jointly with J. Sterbenz pushed this to the energy critical case in $n = 4$ dimensions in [16].

The present paper will borrow quite heavily from [16], and in fact be built directly on the spaces and null-form estimates established there. However, the geometry for the Yang-Mills system is significantly more complicated than for the Maxwell-Klein-Gordon system, as the field A no longer 'essentially behaves like a free wave'. An adaptation of the method of [23] to global regularity for small critical data of high dimensional ($n \geq 6$) Yang-Mills was accomplished in Krieger-Sterbenz [15]. In the present paper we use an approximate parametrix of the same type as in [15]. However, in its construction we take advantage of the better functional framework in [16], as well as of better connection integration techniques borrowed from Wave-Maps [37].

The small data result in the present paper can also be viewed as a stepping stone toward the corresponding large data problem, which is still open. The large data problem is better understood for the Wave-Map equation, where the so-called Threshold Conjecture was recently proved by Sterbenz-Tataru [28, 29], [] and also, independently, by Krieger-Schlag [14] and Tao [31, 32, 33, 34, 35] for special target manifolds. More recently, large data well-posedness was also established for the Maxwell-Klein-Gordon system, independently in Oh-Tataru [20, 21, 22] and Krieger-Luhrman [13].

In related developments, one should also note the work of Bejenaru-Herr [1],[2] on the closely related cubic Dirac equation, as well as the massive Dirac-Klein-Gordon system.

1.2. **Ingredients of the proof.** The present paper is built directly on the predecessor paper [16]. The nonlinearity is split into two parts, a perturbative one and a non-perturbative paradifferential type component. As in [16], even the "perturbative" part cannot directly estimated in full. Instead, there is a portion of it that requires reiteration of the equation and the use of the second null condition. The nonperturbative part is then eliminated via a paradifferential gauge renormalization.

The main novelty here then concerns the approximate parametrix construction for the magnetic potential wave equation (6.1), which is considerably more difficult in the present noncommutative setting. We use an ansatz (6.16) as in [15], but construct the phase shift $O(t, x, \xi)$ via a continuous version of the 'discretized' (over frequency blocks) Gauge construction in [30], see (6.14). Such a construction was first introduced in [37], and proved its usefulness further in [28]. The fact that the angular separation in the definition of the Ψ_k can be chosen as $2^{\delta k}$ with $\delta > 0$ arbitrarily small simplifies the arguments for the control of the parametrix in section 7 compared to the arguments in [15].

1.3. **Notation and Conventions.** We use the notation $A \lesssim B$ to mean $A \leq CB$ for some universal constant $C > 0$. We write $A \ll B$ if the implicit constant should be regarded as small.

Our convention regarding indices is as follows. The greek indices α, β run over $0, \dots, 4$, whereas the latin indices i, j only run over the spatial indices $1, \dots, 4$. We raise and lower indices using the Minkowski metric, and sum over repeated upper and lower indices. The indices k, h, l are reserved for dyadic frequencies.

For the space-time Fourier variables we will use (τ, ξ) or (σ, η). On occasion we set $\tau = \xi_0$, or $\sigma = \eta_0$; we will do this only to keep the notations simple where there is covariant summation with respect to indices α, β.

Littlewood-Paley projections. We denote by $P_k = P_k(D_x)$ the standard spatial Littlewood Paley projections, where k is a dyadic index. We allow k to be either discrete (integer) or continuous. We also use the notations $P_{<k}$, $P_{>k}$ for projections selecting lower or higher frequencies.

On occasion we will also need space-time Littlewood Paley projections. These are denoted by $S_k := S_k(D_{x,t})$, $S_{<k}$, $S_{>k}$.

We also define modulation Littlewood Paley projections, $Q_j := Q_j(|D_t| - |D_x|)$. Sometimes we will restrict these to positive or negative time frequencies, $Q_j^{\pm} := Q^{\pm} Q_j$, where $Q_{\pm} := \mathcal{F}^{-1}[1_{[0,\infty)}(\pm\tau)\mathcal{F}[\varphi]]$ restricts to the $\pm$ frequency half-space.

Frequency envelopes. For some more accurate bounds at various places we need to keep better track of the dyadic frequency distribution of norms. This is done using the language of frequency envelopes. An *admissible frequency envelope* will be any sequence $\{c_k\}_{k\in\mathbb{Z}}$ of positive numbers which is slowly varying upwards,

$$2^{-C_0(j-k)} \leq c_j/c_k \leq 2^{\delta_0(j-k)}, \qquad j > k,$$

with a large universal constant C_0 and a small universal constant δ_0. Given such a sequence and a norm X, we define the norm

$$\|\phi\|_{X_c} = \sup_k c_k^{-1} \|P_k \phi\|_X.$$

We say that c is a frequency envelope for the data $A_x[0]$ if for every $k \in \mathbb{Z}$, we have

$$\|(P_k A_x[0], P_k \phi[0])\|_{\mathcal{H}} \leq c_k.$$

Given any $A_x[0], \phi[0] \in \dot{H}^1 \times L^2$, we may construct such a c by

$$c_k := \sum_{k'>k} 2^{-\delta_0|k-k'|} \|P_{k'} A_x\|_{\mathcal{H}} + \sum_{k'\leq k} 2^{-C_0|k-k'|} \|P_{k'} A_x\|_{\mathcal{H}}.$$

By Young's inequality, we have $\|c\|_{\ell^2} \lesssim \|A_x[0]\|_{\dot{H}^1 \times L^2}$.

Lie group and algebra notations. We use the notation $ad(A)B = [A, B]$ for the Lie bracket on $\mathfrak{g}$, and its interpretation as a representation of $\mathfrak{g}$ as a subspace of $Aut(\mathfrak{g})$. The Killing form

$$\langle A, B \rangle = tr(ad(A)ad(B))$$

is nondegenerate if $\mathbf{G}$ is semisimple, and (with a possible sign adjustment) can be used as an invariant inner product on $\mathfrak{g}$. It also has the invariance property

$$\langle [A, B], C \rangle = \langle A, [B, C] \rangle.$$

The action of $\mathbf{G}$ on g is denoted by $Ad(O)A = OAO^{-1}$. This preserves Lie brackets and the Killing form.

We also need to work with $\mathbf{G}$ valued functions and symbols $O(t,x,\xi)$. To differentiate O we introduce the notations

$$O_{;x} = \partial_x O O^{-1}, \qquad O_{;\xi} = \partial_\xi O O^{-1}, etc.$$

These are all well defined elements of the Lie algebra $\mathfrak{g}$. Furthermore, for any two such derivatives we have the commutation relation

$$\partial_k O_{;l} - \partial_l O_{;k} = [O_{;k}, O_{;l}]. \tag{1.13}$$

Now we introduce the corresponding classes of pseudodifferential operators acting on Lie algebra valued functions. We begin with Lie algebra valued symbols $\Psi(x,\xi)$, where for $\mathfrak{g}$ valued functions B we use the Lie bracket to define using the left calculus

$$Op(ad(\Psi))(x,D)B(x) = \int e^{i(x-y)\xi}[\Psi(x,\xi), B(y)]dyd\xi. \tag{1.14}$$

We note that its L^2 adjoint (with respect to the Killing form duality) is $-Op(ad(\Psi))(D,y)$,

Similarly for a $\mathbf{G}$ valued symbol O we define

$$Op(Ad(O))(x,D)B(x) = \int e^{i(x-y)\xi}O(x,\xi)B(y)O^{-1}(x,\xi)dyd\xi. \tag{1.15}$$

Its L^2 adjoint (with respect to the Killing form duality) is $Op(Ad(O^{-1}))(D,y)$.

1.4. **Structure of the paper.** Our paper is organized as follows:

In Section 2, we begin with some elliptic gauge related fixed time estimates. In particular these will help us relate the full nonlinear gauge independent energy with the linear energy associated to the MKG-CG system. We also consider similar issues for the linearized equation.

In the following section we switch to space-time analysis, and define the function spaces S^1 and N; with minor changes this follows [16]. We also recall some useful estimates from [16], and add to that some additional properties from [21], related to the interval decomposition of the S^1 and N spaces.

In Section 4 we use the S^1 norms to provide a stronger form of our main theorem, and we show that this follows from three estimates in Propositions 4.1, 4.2 and 4.3.

Section 5 contains the perturbative part of our analysis, which primarily consists of bilinear estimates in S^1 and N spaces. There we prove Proposition 4.1, as well as Proposition 4.3 (the latter modulo Lemma 5.6, which captures the trilinear structure governed by the second null condition, and whose proof is relegated to the next to last section).

The bulk of the paper is devoted to the construction of a parametrix for the paradifferential equation (4.3), which is the main step in the proof of the remaining Proposition 4.2.

We begin in Section 6 with some heuristic considerations, followed by the rigorous definition of the parametrix and by Theorem 3, which summarizes its properties. This suffices for the proof of Proposition 4.2. In Section 7 we review the notion of decomposability, and establish a number of bounds for the symbols Ψ and O arising in the definition of the parametrix. The symbol bounds are then used in Section 8 to derive kernel bounds, and a number of L^2 estimates, concluding with the proof of the first three parametrix bounds in Theorem 3, as well as the Strichartz and null frame bounds for the renormalization operators

in our parametrix. Section 9 contains the proof of the error estimates in Theorem 3, modulo Lemma 9.1. The two estimates that require a fine trilinear analysis, namely Lemma 9.1 and Lemma 5.6, are proved in Section 10.

2. Elliptic L^2 bounds

Here for convenience we show that any small energy data admits a Coulomb representation which is small in $\mathcal{H}$. We also show that the equations (1.7)-(1.8) are well-posed; this justifies the fact that the initial data in the Coulomb gauge is fully determined by $(A_j(0), \partial_t A_j(0))$ (at least at small energies).

Proposition 2.1. *a) Let $(\tilde{A}_\alpha(0), \partial_t \tilde{A}_j(0)) \in \dot{H}^1 \times L^2$ be an initial data for the Yang-Mills equation with energy E. If E is small enough then there exists an unique gauge equivalent Coulomb data with*

$$\|(A_j(0), \partial_t A_j(0))\|_{\mathcal{H}}^2 \approx E \tag{2.1}$$

b) For any Coulomb data $(A_j(0), \partial_t A_j(0))$ which is small in $\mathcal{H}$ there exists a unique solution $(A_0(0), \partial_t A_0(0)) \in \mathcal{H}$ to (1.7)-(1.8) *so that*

$$\|(A_0(0), \partial_t A_0(0))\|_{\mathcal{H}}^2 \lesssim E^2 \tag{2.2}$$

c) If in addition we have $(A_j(0), \partial_t A_j(0)) \in \mathcal{H}^N$ then we also have $(A_0(0), \partial_t A_0(0)) \in \mathcal{H}^N$ and

$$\|(A_0(0), \partial_t A_0(0))\|_{\mathcal{H}^N}^2 \lesssim E\|(A_j(0), \partial_t A_j(0))\|_{\mathcal{H}^N}^2 \tag{2.3}$$

Proof. The first part is proved (in $n \geq 6$ dimensions, but equally valid in lower ones) for example in [15]. The second part is a consequence of Sobolev embeddings and a simple fixed point argument.

□

We also consider the counterpart of part (b) for the linearized equation (1.10). We have:

Proposition 2.2. *Let $(A_j(0), \partial_t A_j(0)) \in \mathcal{H}$ be a Coulomb initial data for the Yang-Mills equation with small energy E. Let $\frac{1}{2} < s < 1$ and $(B_j(0), \partial_t B_j(0)) \in \dot{\mathcal{H}}^s$ be a Coulomb initial data for the linearized Yang-Mills equation* (1.10)*. Then there exists a unique solution $(B_0(0), \partial_t B_0(0)) \in \dot{\mathcal{H}}^s$ to* (1.11)-(1.12) *so that*

$$\|(B_0(0), \partial_t B_0(0))\|_{\dot{\mathcal{H}}^s}^2 \lesssim E\|(B_j(0), \partial_t B_j(0))\|_{\dot{\mathcal{H}}^s}^2 \tag{2.4}$$

Proof. This is also a simple fixed point argument which is based on the Sobolev embeddings. The details are left for the reader.

□

3. The S and N spaces

With minor modifications, we will use the function spaces introduced in [16] in the whole of $\mathbb{R}^{4+1}$. We also need to work on bounded time intervals, for which we use the set-up of [21].

3.1. **The S^1, N, Z and Y^1 spaces.** We begin our discussion with the function spaces introduced in [16], namely S^1 for the MKG waves (A,ϕ) and N for the inhomogeneous terms in both the $\Box$ and the $\Box_A$ equation. In addition to these we also recall the Z norm, which plays a key role in the reiteration of the equation in connection to trilinear estimates and the second null structure.

These are spaces of functions defined over all of $\mathbb{R}^{n+1}$, together with the related spaces S and N^*. They are all defined via their dyadic subspaces, with norms

$$\|\phi\|_X^2 = \sum_{k\in\mathbb{Z}} \|\phi_k\|_{X_k}^2, \qquad X \in \{S, S^1, N, Z\}.$$

Here we use the ℓ^2 Besov structure. On occasion we will also need ℓ^1 and ℓ^∞ type Besov norms, which are denoted by $\ell^1 X$, respectively $\ell^\infty X$, with norms

$$\|\phi\|_{\ell^1 X} = \sum_{k\in\mathbb{Z}} \|\phi_k\|_{X_k}, \qquad \|\phi\|_{\ell^\infty X} = \sup_{k\in\mathbb{Z}} \|\phi_k\|_{X_k}, \qquad X \in \{S, S^1, N, Z\}.$$

We recall the definition of their norms. With minor modifications at high modulations, we follow [16]. For N_k we set

$$N_k \ = \ L^1L^2 + X_1^{0,-\frac{1}{2}}, \tag{3.1}$$

where

$$\|\phi\|_{X_r^{s,b}} := \Big(\sum_k \Big(\sum_j (2^{sk}2^{bj}\|P_kQ_j\phi\|_{L^2L^2})^r\Big)^{\frac{2}{r}}\Big)^{\frac{1}{2}}.$$

The N_k norm is the same as in [16].

The S_k space is a strengthened version of N_k^*,

$$X_1^{0,\frac{1}{2}} \subseteq S_k \subseteq L^\infty L^2 \cap X_\infty^{0,\frac{1}{2}} = N_k^*, \tag{3.2}$$

while S_k^1 is defined as

$$\|\phi\|_{S_k^1} = \|\nabla\phi\|_{S_k} + 2^{-\frac{k}{2}}\|\Box\phi\|_{L^2L^2} + 2^{-\frac{4k}{9}}\|\Box\phi\|_{L^{\frac{9}{5}}L^2}. \tag{3.3}$$

As in [21], compared to [16] we have loosened the ℓ^1 summability of the $\Box^{-1}L^2L^2$ norm and added the $\Box^{-1}L^{\frac{9}{5}}L^2$ norm above. Both of these modifications are of interest only at high modulations. The exact exponent $9/5$ is not really important, for our purposes it only matters that it is less than two and greater than $5/3$.

We now recall the definition of the space S_k from [16]. The space S_k scales like free waves with $L^2 \times \dot{H}^{-1}$ initial data, and is defined by

$$\|\phi\|_{S_k}^2 \ = \ \|\phi\|_{S_k^{str}}^2 + \|\phi\|_{S_k^{ang}}^2 + \|\phi\|_{X_\infty^{0,\frac{1}{2}}}^2 \ ,$$

where:

$$\begin{aligned} \|\phi\|_{S_k^{str}} \ &= \sup_{2\leq q,r,\leq\infty,\ \frac{1}{q}+\frac{3/2}{r}\leq\frac{3}{4}} 2^{(\frac{1}{q}+\frac{4}{r}-2)k}\|(\phi, 2^{-k}\partial_t\phi)\|_{L^qL^r}\ , \quad \|\phi\|_{S_k^{ang}} = \sup_{l<0}\|\phi\|_{S_{k,k+2l}^{ang}}\ , \\ \|\phi\|_{S_{k,j}^{ang}}^2 &= \sum_\omega \|P_l^\omega Q_{<k+2l}\phi\|_{S_k^\omega(l)}^2 \qquad \text{with } l = \lceil \frac{j-k}{2} \rceil. \end{aligned} \tag{3.4}$$

The S_k^{str} norm controls all admissible Strichartz norms on $\mathbb{R}^{1+4}$. The ω-sum in the definition of $S_{k,j}^{ang}$ is over a covering of $\mathbb{S}^3$ by caps ω of diameter 2^l with uniformly finite overlaps,

and the symbols of P_l^ω form a smooth partition of unity associated to this covering. The angular sector norm $S_k^\omega(l)$ combines the null frame space as in wave maps [30, 36] with additional square-summed norms over smaller radially directed blocks $\mathcal{C}_{k'}(l')$ of dimensions $2^{k'} \times (2^{k'+l'})^3$. We first define

$$\begin{aligned} \|\phi\|_{PW_\omega^\pm(l)} &= \inf_{\phi=\int\phi^{\omega'}} \int_{|\omega-\omega'|\leqslant 2^l} \|\phi^{\omega'}\|_{L^2_{\pm\omega'}(L^\infty_{(\pm\omega')^\perp})} d\omega' \ , \\ \|\phi\|_{NE} &= \sup_\omega \|\not{\nabla}_\omega \phi\|_{L^\infty_\omega(L^2_{\omega^\perp})} \ , \end{aligned}$$

where the norms are with respect to $\ell_\omega^\pm = t \pm \omega\cdot x$ and the transverse variable in the $(\ell_\omega^\pm)^\perp$ hyperplane (i.e., constant $\ell_\omega^\pm$ hyperplanes). Moreover, $\not{\nabla}_\omega$ denotes tangential derivatives on the $(\ell_\omega^+)^\perp$ hyperplane. As in [16], we set:

$$\begin{aligned} \text{(3.5)} \quad \|\phi\|^2_{S_k^\omega(l)} &= \|\phi\|^2_{S_k^{str}} + 2^{-2k}\|\phi\|^2_{NE} + 2^{-3k}\sum_\pm \|Q^\pm\phi\|^2_{PW_\omega^\mp(l)} \\ &+ \sup_{\substack{k'\leqslant k, l'\leqslant 0 \\ k+2l\leqslant k'+l'\leqslant k+l}} \sum_{\mathcal{C}_{k'}(l')} \Big(\|P_{\mathcal{C}_{k'}(l')}\phi\|^2_{S_k^{str}} + 2^{-2k}\|P_{\mathcal{C}_{k'}(l')}\phi\|^2_{NE} \\ &\qquad + 2^{-2k'-k}\|P_{\mathcal{C}_{k'}(l')}\phi\|^2_{L^2(L^\infty)} + 2^{-3(k'+l')}\sum_\pm \|Q^\pm P_{\mathcal{C}_{k'}(l')}\phi\|^2_{PW_\omega^\mp(l)} \Big) \ , \end{aligned}$$

where the $\mathcal{C}_{k'}(l')$ sum runs over a covering of $\mathbb{R}^4$ by the blocks $\mathcal{C}_{k'}(l')$ with uniformly finite overlaps, and the symbols of $P_{\mathcal{C}_{k'}(l')}$ form an associated partition of unity. We emphasize the role played by the next to last term in the above expression, which captures the gain in Strichartz estimates on blocks which are shorter radially. This gain was first discovered in [11], and plays a key role in getting some of the sharper bilinear bounds which are needed in the present paper. We remark that there is a similar gain at the level of the L^2L^6 Strichartz norm, which could be easily added to the S^1 structure; this would improve some of the intermediate estimates in this paper, but would not affect the final result in a significant way.

We also define the smaller space $S_k^\sharp \subset S_k$ (see the bound (3.7) below) by

$$\|u\|_{S_k^\sharp} = \|\Box u\|_{N_k} + \|\nabla u\|_{L^\infty L^2}.$$

On occasion we need to separate the two characteristic cones $\{\tau = \pm|\xi|\}$. Thus we define the spaces $N_{k,\pm}$, $S_{k,\pm}^\sharp$ and $N_{k,\pm}^*$ in an obvious fashion, so that

$$N_k = N_{k,+} \cap N_{k,-}, \quad S_k^\sharp = S_{k,+}^\sharp + S_{k,-}^\sharp, \quad N_k^* = N_{k,+}^* + N_{k,-}^* \ .$$

Next we describe an auxiliary space of the type $L^1(L^\infty)$ which will be useful for decomposing the nonlinearity:

$$\|\phi\|_Z^2 = \sum_k \|P_k\phi\|^2_{Z_k} \ , \quad \|\phi\|^2_{Z_k} = \sup_{l<C}\sum_\omega 2^l \|P_l^\omega Q_{k+2l}\phi\|^2_{L^1(L^\infty)} \ .$$

Note that as defined this space already scales like $\dot{H}^1$ free waves. In addition, note the following useful embedding which is a direct consequence of Bernstein's inequality:

$$\text{(3.6)} \qquad \Box^{-1}L^1(L^2) \subseteq Z \ .$$

Finally, the function space Y_1 for A_0 is simple to describe, since the A_0 equation is elliptic:

$$\|A_0\|_{Y^1}^2 = \|\nabla A_0\|_{L^\infty L^2}^2 + \|\nabla A_0\|_{L^2 \dot{H}^{\frac{1}{2}}}^2 ,$$

One of the results in [16] asserts that we have linear solvability for the d'Alembertian in our setting.

Proposition 3.1. *We have the linear estimates*

$$\|\nabla\phi\|_S \lesssim \|\phi[0]\|_{\mathcal{H}} + \|\Box\phi\|_N , \tag{3.7}$$

$$\|\phi\|_{S^1} \lesssim \|\phi[0]\|_{\mathcal{H}} + \|\Box\phi\|_{N \cap L^2 \dot{H}^{-\frac{1}{2}} \cap L^{\frac{9}{5}} \dot{H}^{-\frac{4}{9}}} . \tag{3.8}$$

Here (3.7) is the embedding $S^\sharp \subset S$, whereas (3.8) follows immediately from (3.7).

3.2. **Interval localization.** So far, we have described the global setting in [16]. However, in this article we need to work on compact time intervals, therefore we also need suitable interval localized function spaces. For this we borrow the set-up of [21].

We start by defining

$$\|\phi\|_{S^1[I]} = \inf_{\phi = \widetilde{\phi}_{|I}} \|\widetilde{\phi}\|_{S^1}, \qquad \|f\|_{N[I]} = \inf_{f = \tilde{f}_{|I}} \|\tilde{f}\|_N \tag{3.9}$$

The next result from [21] provides an alternate take on these definitions:

Proposition 3.2. *Consider a time interval $I = [0,T]$, and its characteristic function χ_I. Then we have the bounds*

$$\|\chi_I \phi\|_S \lesssim \|\phi\|_S, \qquad \|\chi_I f\|_N \lesssim \|f\|_N, \tag{3.10}$$

The latter norm is also continuous as a function of I. We also have the linear estimates

$$\|\nabla\phi\|_{S[I]} \lesssim \|\phi[0]\|_{\mathcal{H}} + \|\Box\phi\|_{N[I]}, \tag{3.11}$$

$$\|\phi\|_{S^1[I]} \lesssim \|\phi[0]\|_{\mathcal{H}} + \|\Box\phi\|_{(N \cap L^2 \dot{H}^{-\frac{1}{2}} \cap L^{\frac{9}{5}} \dot{H}^{-\frac{4}{9}})[I]}. \tag{3.12}$$

Note that a consequence of the above proposition is that, up to equivalent norms, we can replace the arbitrary extensions in (3.9) by the zero extension in the N case, respectively by homogeneous waves with $(\phi, \partial_t \phi)$ as the data at each endpoint outside I in the S^1 case.

4. The proof of the main result

In this section we provide the main intermediate results used in the proof, and we use them in order to complete the proof of the Theorem 1. For convenience, we restate the theorem here in a more precise form:

Theorem 2. *The Yang-Mills system in Coulomb gauge* (1.6)-(1.7)-(1.8) *is globally well-posed in $\dot{H}^1 \times L^2$ for initial data which is small in $\mathcal{H} = \dot{H}^1 \times L^2$,*

$$\|A_x(0), \partial_t A_x(0)\|_{\mathcal{H}} \le \epsilon, \tag{4.1}$$

in the following sense:

(i) (Regular data) If in addition the data (A_{0j}, A_{1j}) is more regular, $(A_{0j}, A_{1j}) \in \mathcal{H}^N$, then there exists a unique global in time regular solution $(A_j, \partial_0 A_j) \in S^N$, which has a Lipschitz dependence on the initial data locally in time in the $\mathcal{H}^N$ topology.

(ii) (Rough data) The initial data to solution map admits an extension

$$\mathcal{H} \ni (A_{j0}, A_{j1}) \to (A_j, \partial_t A_j) \in S^1,$$

globally in time, for all small data as above, and which is continuous in the $\mathcal{H}\cap\dot{H}^s \to S^1\cap\dot{S}^s$ *topology (applied to differences of solutions) for* $s<1$ *but close to* 1.

To set the stage for the proof of the theorem, we assume that we have a solution A_j for the Yang mills equation (1.6) in a time interval I containing 0, and further that this solution satisfies

$$\|A_j\|_{S^1[I]} \le \epsilon \ll 1. \tag{4.2}$$

We begin by rewriting the equation in a paradifferential fashion,

$$\Box A_{j,k} + 2\mathbf{P}[A_{\alpha,<k}, \partial^\alpha A_{j,k}] = F_k, \tag{4.3}$$

where F_k contains only terms that will be treated in a perturbative fashion,

$$F_k = \mathbf{P}\left(P_k\left([A^\alpha,\partial_j A_\alpha] - 2[A^\alpha_{\ge k},\partial_\alpha A_j] - [\partial_0 A_0, A_j] - [A^\alpha,[A_\alpha,A_j]]\right) - 2[[P_k, A^\alpha_{<k}],\partial_\alpha A_j]\right). \tag{4.4}$$

To estimate F we use the following:

Proposition 4.1. *Assume that* A *is a solution to the Yang-Mills equation in Coulomb gauge in an interval* I*, which satisfies* (4.2)*. Then for any admissible frequency envelope* c *we have*

$$\|F\|_{N_{c^2}[I]} \lesssim \|A_j\|^2_{S^1_c[I]}. \tag{4.5}$$

This proposition is proved in the next section. We remark that by appropriately choosing the envelope c, this implies that

$$\|F\|_{N_c[I]} \lesssim \epsilon\|A_j\|_{S^1_c[I]}, \tag{4.6}$$

as well as

$$\|F\|_{\ell^1 N[I]} \lesssim \epsilon\|A_j\|_{S^1[I]}, \tag{4.7}$$

We now turn our attention to the linear equation (4.3). In order to uncouple variables it will be useful to also consider the more general frequency localized equation:

$$\Box B_{j,k} + 2\mathbf{P}[A_{\alpha,<k}, \partial^\alpha B_{j,k}] = G_{j,k}. \tag{4.8}$$

Proposition 4.2. *Assume that* A *is a solution to the Yang-Mills equation in Coulomb gauge in an interval* I*, which satisfies* (4.2)*. Then for the equation* (4.8) *we have the following linear estimate:*

$$\|B_{j,k}\|_{S^1[I]} \lesssim (\|G_{j,k}\|_{N[I]} + \|B_{j,k}[0]\|_{\mathcal{H}}). \tag{4.9}$$

This result is the key point of the paper. Its proof is closed in Section 6, using the paradifferential parametrix in Theorem 3. However, the proof of Theorem 3 requires all the subsequent sections of the paper.

The two bounds above suffice in order to close the a-priori bounds in S^1 and S^N, including frequency envelope bounds. In order to compare different solutions, we need to work with the linearized equation (1.10)-(1.11)-(1.12).

Proposition 4.3. *Suppose that* A *is a solution to the Yang-Mills equation in Coulomb gauge in an interval* I*, which satisfies* (4.2)*. Then the equation* (1.10) *is well-posed in* $\mathcal{H}^s$ *for* $s<1$*, close to* 1*, in the time interval* I*.*

To further clarify this last result, we rewrite the equation (1.10) in a paradifferential form,

$$\Box B_k + \mathbf{P}[A_{\alpha,<k}\partial^\alpha B_k] = \mathbf{P}[B_{\alpha,<k}, \partial^\alpha A_k] + G_k. \tag{4.10}$$

The term G_k plays the same role as F_k in the original equation. Precisely, we have:

Proposition 4.4. *Assume that A is a solution to the Yang-Mills equation in Coulomb gauge in an interval I, which satisfies* (4.2). *Then for $s \leq 1$, close to* 1 *we have*

$$\|G_j\|_{N^{s-1}} \lesssim \epsilon \|B_j\|_{S^s}. \tag{4.11}$$

This result is proved in the next section. We remark that the range of s depends on the constant δ in the estimate (5.1) in the next section, which is from [16]. We expect the correct range here to be $s > \frac{1}{2}$.

The new term $[B_{\alpha,<k}, \partial^\alpha A_k]$ in (4.10) does not have a counterpart in the previous argument. This is the term which is responsible for disallowing case $s = 1$ in Proposition 4.2, and ultimately for the failure of the Lipschitz dependence of the solution on the initial data in the strong topology $\mathcal{H}$. We will estimate this in a more roundabout fashion, proving the following statement:

Proposition 4.5. *Suppose that $B \in S^s$ solves the linearized equation* (1.10) *in a time interval I, around a YM-CG solution A which satisfies* (4.2). *Then for $s < 1$, close to* 1 *we have the estimate*

$$\|[B_{\alpha,<k}, \partial^\alpha C_k]\|_{N^{s-1}} \lesssim \epsilon \|B\|_{S^s} \|C_k\|_{S^1}. \tag{4.12}$$

This Proposition is more delicate than the previous proposition, as it requires a fine trilinear analysis based on reiterating the linearized equation. Its proof is also in the next section, modulo the most difficult case in Lemma 5.6, which is relegated to Section 10.

The result in Proposition 4.3 is a direct consequence of Proposition 4.2, Proposition 4.4 and Proposition 4.5. We now turn our attention to Theorem 2.

Proof of Theorem 2. Here we show that Theorem 2 follows from Propositions 4.14.2,4.3. In addition to these Propositions, we will also take it for granted that for large N (e.g. $N \geq 3$) the Yang-Mills equation is locally well-posed in $\mathcal{H}^N$, with smooth dependence on the initial data; at least at small energies this is a straightforward perturbative result, based purely on energy estimates. We carry this out in several steps.

Step 1:(*A-priori bounds for regular data*) Here we consider regular $\mathcal{H}^N$ solutions in a time interval $I = [0, T]$, and which satisfy the smallness condition

$$\|A_x\|_{S^1[I]} \leq \epsilon_0 \ll 1. \tag{4.13}$$

Let c be an admissible $\mathcal{H}$ frequency envelope for the initial data. Then we claim that c is also an S^1 frequency envelope for the solution, and, in addition, we have the bound

$$\|A_x\|_{S^1_c} \lesssim \|A_x[0]\|_{\mathcal{H}_c}. \tag{4.14}$$

We remark that, as a consequence of this, we have in particular the bounds

$$\|A_x\|_{S^1} \lesssim \|A_x[0]\|_{\mathcal{H}}, \qquad \|A_x\|_{S^N} \lesssim \|A_x[0]\|_{\mathcal{H}^N}. \tag{4.15}$$

Assume first that we already know that $A_x \in S_c$. Then (4.14) is obtained by successively applying Propositions 4.1, 4.2 in the equation (4.3). Without knowing that $A_x \in S_c$, let d be

an admissible frequency envelope for A_x in S^1. Then for $\delta > 0$ we have $A_x \in S^1_{c+\delta d}$. Then we have (4.14) with c replaced by $c + \delta d$, and it suffices to let δ to zero to obtain again (4.14).

Step 2:(*Global solutions for regular data*) Here we start with regular data $(A_j(0)\, \partial_t A_j(0)) \in \mathcal{H}^N$ which is small in the energy norm, i.e. it satisfies (4.1). Then the solution exists in $\mathcal{H}^N$ on some nonempty time interval $[0, T)$. We claim that the solution is global, $T = \infty$, and that it satisfies the bound

$$\|A_j\|_{S^1} \leq C\epsilon, \tag{4.16}$$

with a fixed universal constant C.

This is done using a time continuity argument. Let $\mathcal{T}$ denote the set of all times T for which a classical (i.e. $\mathcal{H}^N$ solution) exists in $[0, T]$ which satisfies (4.16). We will prove that $\mathcal{T}$ is both open and closed, and thus must be equal to $\mathbb{R}^+$.

a) $\mathcal{T}$ is closed. Indeed, suppose that $[0, T_0) \subset \mathcal{T}$. By (4.15) we have a uniform bound

$$\|A_j\|_{S^N[0,T]} \lesssim \|A_j[0]\|_{\mathcal{H}^N}.$$

Then, in view of the Lipschitz dependence for classical solutions, the solution A_j extends to time T_0 (and indeed, past it) as a classical solution. By a scaling argument, see e.g. [37], the $S^1[I]$ norm of classical solutions depends continuously on the interval I. Thus the bound (4.16) at time T_0 follows, so $T_0 \in \mathcal{T}$.

b) $\mathcal{T}$ is open. Let $T \in \mathcal{T}$. Then $A_j[T] \in H^N$, so we can continue the solution beyond time T. It remains to show that the bound (4.16) persists. Using again the continuous dependence of the $S^1[I]$ norm of classical solutions on the interval I, it suffices to prove (4.16) this under a bootstrap assumption

$$\|A_j\|_{S^1} \leq 2C\epsilon, \tag{4.17}$$

with a large universal constant C. But this again follows from (4.15) in Step 1.

Step 3:(*Weak Lipschitz dependence for regular solutions*) Here we assert that for any two small data global regular solutions we have the bound

$$\|A_j - \tilde{A}_j\|_{S^s} \lesssim \|A_j[0] - \tilde{A}_j[0]\|_{\mathcal{H}^s}. \tag{4.18}$$

provided $s < 1$ is close to 1. This is a direct consequence of the result in Proposition 4.3.

Step 4:(*Rough data solutions*) The continuous extension of the flow map to rough data for solutions which satisfy (4.16), using the $\mathcal{H} \cap \dot{H}^s$ topology, follows in a standard manner from two properties of small data solutions:

- The frequency envelope bounds (4.14).
- The Lipschitz dependence in a weaker topology (4.18).

Indeed, consider some small energy data $A_x[0] \in \mathcal{H}$. Then for any $A_x^{(n)}$ are regular solutions, whose data $A_x^{(n)}[0]$ converge to $A_x[0] \in \mathcal{H}$ in the sense that

$$\|A_x^{(n)}[0] - A_x[0]\|_{\mathcal{H} \cap \dot{H}^s} \to 0.$$

By (4.18) the limit A_x of $A_x^{(n)}$ exists in $\dot{S}^s$. Further, the relation (4.18) extends to all solutions constructed in this way.

Favorably choosing $A_x^{(n)}[0]$ so that they have the same $\mathcal{H}$ frequency envelope as $A_x[0]$ (e,g. as $A_x^{(n)}[0] = P_{<n} A_x[0]$) and applying (4.14), it follows that $A_x \in S^1$, and further that (4.14) holds for A_x.

Finally, to establish the continuity of the data to solution map from $\mathcal{H} \cap \dot{\mathcal{H}}^s$ to $S \cap \dot{S}^s$ we use the previously established $\dot{H}^s$ Lipschitz bound for low frequencies, combined with the uniform smallness of high frequency tails, which is in turn derived from the frequency envelope bound.

□

5. Bilinear estimates and perturbative analysis

The first goal of this section is to review the bilinear null form bounds from [16], which will be repeatedly used in our analysis. Then we use these bounds to provide some preliminary characterization of YM solutions which satisfy an a-priori S^1 bound. Finally, we conclude with a proof of Propositions 4.1 and 4.3.

5.1. **Bilinear null form bounds.** We begin with the main bilinear null form estimate, where $\mathcal{N}(u, v)$ refers to any expression of the form $\partial_i u \partial_j v - \partial_j u \partial_i v$. It comes from [16], and specifically from (131) in Theorem 12.1 there:

Proposition 5.1. *([16]) For any null form $\mathcal{N}$ we have the following null form estimates:*

$$\|P_k \mathcal{N}(u_{k_1}, v_{k_2})\|_N \lesssim 2^k 2^{\delta(k_{min} - k_{max})} \|u_{k_1}\|_S \|v_{k_2}\|_S \tag{5.1}$$

We remark that, in view of Proposition 3.2, the same bound holds in any time interval I.

Ideally, we would like to improve this bound in the case of low-high frequency interactions $k_1 < k_2 = k$, and have a 2^{k_1} factor instead. Unfortunately that does not work in general. However, it does work for the most part. To describe that we isolate the bad component, namely $\mathcal{H}^* \mathcal{N}(u_{k_1}, v_{k_2})$. Here, following [16], if $\mathcal{M}(D_{t,x}, D_{t,y})$ is any bilinear translation invariant operator then we set:

$$\mathcal{H}^* \mathcal{M}(\phi_{k_1}, \psi_{k_2}) \ = \ \sum_{j<k_1} Q_{<j-C} \mathcal{M}(Q_j \phi_{k_1}, Q_{<j-C} \psi_{k_2}) \ , \qquad k_1 < k_2 - C \tag{5.2}$$

We observe that the map $\mathcal{H}^*$ selects the portion of the bilinear interaction where both the high frequency input and the output have high modulation. This case is unfavorable in the high frequency limit; this is most easily seen using duality to rewrite the above bound in a trilinear fashion. We also remark that the frequency/modulation localization in $\mathcal{H}^*$ fixes the angle θ between the two input functions to

$$\theta \approx 2^{(j-k_1)/2}$$

A benefit of the null form structure of the nonlinearity is that it provides an additional gain at small angles in bilinear estimates, which is roughly proportional to the angle. We will also need to take advantage of this gain in our estimates. For this we introduce a second selection device for bilinear interactions. Precisely, given two spatial frequencies ξ and η we define a partition of unity

$$1 = \sum_{\theta \text{ dyadic}} \chi_\theta(\xi, \eta)$$

where $\chi_\theta(\xi,\eta)$ is a smooth homogeneous cutoff which selects the region where $\angle(\xi,\eta)\approx\theta$. Then, given bilinear translation invariant operator $\mathcal{M}(D_{t,x},D_{t,y})$ with symbol $m(\tau,\xi,\sigma,\eta)$, we define $\mathcal{M}^\theta$ as the bilinear translation invariant operator with symbol $m(\tau,\xi,\sigma,\eta)\chi_\theta(\xi,\eta)$. We will similarly used the notations $\mathcal{M}^{<\theta}$, $\mathcal{M}^{>\theta}$ with the obvious meanings.

We now return to the promised decomposition of the null form into a good and a bad part. For the complement $(I-\mathcal{H}^*)\mathcal{N}(u_{k_1},v_{k_2})$ we have a good S bound; for $\mathcal{H}^*\mathcal{N}(u_{k_1},v_{k_2})$, instead, we use the Z norm as a proxy. The following estimates are contained in Theorem 12.1, Theorem 12.2 in [16]:

Proposition 5.2. *([16]) For $k_1<k_2-C$ and any null form $\mathcal{N}$ we have the following bilinear estimates:*

a) $S^1\times S^1\to N$ bound:

$$\|(I-\mathcal{H}^*)\mathcal{N}(u_{k_1},v_{k_2})\|_N\lesssim 2^{k_1}\|u_{k_1}\|_{S^1}\|u_{k_2}\|_{S^1}. \tag{5.3}$$

We also have the small angle improvement

$$\|(I-\mathcal{H}^*)^{<\theta}\mathcal{N}(u_{k_1},v_{k_2})\|_N\lesssim 2^{k_1}\theta^{\frac14}\|u_{k_1}\|_{S^1}\|u_{k_2}\|_{S^1}. \tag{5.4}$$

b) $Z\times S^1\to N$ bound:

$$\|\mathcal{H}^*\mathcal{N}(u_{k_1},v_{k_2})\|_N\lesssim 2^{k_1}\|u_{k_1}\|_Z\|v_{k_2}\|_{S^1},\qquad k_1<k_2 \tag{5.5}$$

c) $L^2\dot H^{\frac32}\times S\to N$ bound:

$$\|(I-\mathcal{H}^*)(u_{k_1}\cdot\nabla v_{k_2})\|_N\lesssim\|u_{k_1}\|_{L^2\dot H^{\frac32}}\|u_{k_2}\|_{S^1}. \tag{5.6}$$

d) $\Box^{\frac12}\Delta^{-\frac12}Z\times S\to N$ bound:

$$\|\mathcal{H}^*(u_{k_1}\cdot\nabla v_{k_2})\|_N\lesssim\|u_{k_1}\|_{\Box^{\frac12}\Delta^{-\frac12}Z}\|u_{k_2}\|_S. \tag{5.7}$$

In order to be able to take advantage of the bilinear bounds which use the Z norm we need to have an additional estimate allowing us to bound Z norms appropriately.

To describe the result we need a second operator $\mathcal{H}_k$, which, following [16] is defined as

$$\mathcal{H}_k\mathcal{M}(\phi_{k_1},\psi_{k_2}) = \sum_{j<k+C}Q_jP_k\mathcal{M}(Q_{<j-C}\phi_{k_1},Q_{<j-C}\psi_{k_2}),\qquad k<k_1=k_2 \tag{5.8}$$

Then the Z bounds are as follows, also contained in [16]:

Proposition 5.3. *For any null form $\mathcal{N}$ have the following Z bounds:*

a) Bound for classical solutions:

$$\|\phi_k\|_Z\lesssim\|\Box\phi_k\|_{L^1L^2} \tag{5.9}$$

b) High-low interactions:

$$\|P_k\mathcal{N}(u_{k_1},v_{k_2})\|_{\Box Z}\lesssim 2^k2^{-\delta|k_1-k_2|}\|u_{k_1}\|_S\|u_{k_2}\|_S,\qquad k>k_{max}-C \tag{5.10}$$

$$\|P_k(u_{k_1}\cdot\nabla v_{k_2})\|_{\Delta^{\frac12}\Box^{\frac12}Z}\lesssim 2^{k_1+k_2}2^{-\delta|k_1-k_2|}\|u_{k_1}\|_S\|u_{k_2}\|_S,\qquad k>k_{max}-C \tag{5.11}$$

c) High-high-low interactions:

$$\|(I-\mathcal{H}_k)\mathcal{N}(u_{k_1},v_{k_2})\|_{\Box Z}\lesssim 2^{k_1}2^{-\delta|k-k_1|}\|u_{k_1}\|_S\|u_{k_2}\|_S,\qquad k<k_1=k_2 \tag{5.12}$$

$$\|(I-\mathcal{H}_k)(u_{k_1}\cdot\nabla v_{k_2})\|_{\Delta^{\frac12}\Box^{\frac12}Z}\lesssim 2^{k_1+k_2}2^{-\delta|k-k_1|}\|u_{k_1}\|_S\|u_{k_2}\|_S,\qquad k<k_1=k_2 \tag{5.13}$$

To better understand how the last two propositions fit together, we remark that the in the bounds in Proposition 5.2 there is no off-diagonal decay with respect to the frequency gap $k_1 - k_2$. Hence, we can only apply it for portions of A_x which we control in $\ell^1 Z$. This is why the off-diagonal decay in (5.10), (5.11) and (5.12), (5.13) is important.

We further remark that the same estimates in [16] also yield a bound for the remaining bad component of $\mathcal{N}(u_{k_1}, v_{k_2})$, namely

$$\|\mathcal{H}_k \mathcal{N}(u_{k_1}, v_{k_2})\|_{\Box Z} \lesssim 2^{k_1} \|u_{k_1}\|_S \|u_{k_2}\|_S, \qquad k < k_1 = k_2 \tag{5.14}$$

and similarly

$$\|\mathcal{H}_k (u_{k_1} \cdot \nabla v_{k_2})\|_{\Delta^{\frac{1}{2}} \Box^{\frac{1}{2}} Z} \lesssim 2^{k_1+k_2} \|u_{k_1}\|_S \|u_{k_2}\|_S, \qquad k < k_1 = k_2 \tag{5.15}$$

Unfortunately, these bounds have no off-diagonal decay, so they only lead to an $\ell^\infty Z$ bound for the corresponding "bad" part of A. If one attempts to combine this with Proposition 5.2, we are left with an unresolved logarithmic divergence. Addressing this issue requires the finer trilinear analysis in the last section of the paper, and the use of the second null form.

5.2. **Characterization of S^1 solutions for YM-CG.** While the S^1 envelope of a Yang-Mills wave A naturally inherits the ℓ^2 dyadic structure from the initial data, one might expect that the inhomogeneous part of A, arising from bilinear or cubic interactions, might carry a better, ℓ^1 dyadic summation. This was indeed the case for the Maxwell-Klein Gordon system in [16], and it allowed us to treat the inhomogeneous part of A in a perturbative fashion, as well as to use free wave magnetic potentials in the parametrix construction. Unfortunately, it is no longer the case here, as the bilinear self-interactions of A are not perturbative. However, we are still able to prove ℓ^1 dyadic summation fully for A_0, and in a partial manner only for the inhomogeneous part of A_x. This will allow us to treat not all but the bulk of the nonlinearity in a perturbative fashion. Precisely, we prove the following:

Proposition 5.4. *Let A be a solution for the YM-CG in an interval I so that $\|A\|_{S^1} \leq \epsilon$. Then the following property holds:*

$$\|\nabla A_0\|_{\ell^1 L^2 \dot{H}^{\frac{1}{2}}} \lesssim \epsilon. \tag{5.16}$$

Also, for each $0 \leq b < \frac{1}{2}$ we have

$$\|\Box A_x\|_{\ell^1 X^{b-\frac{1}{2}, -b}} \lesssim_\delta \epsilon \tag{5.17}$$

A related result holds for the linearized equation. There, the dyadic summation is not an issue because the bounds for the linearized problem are no longer at scaling (though they are scale invariant). Also, the bounds we need for the linearized equation are not as refined as those we need for the original equation. We have:

Proposition 5.5. *Let A be a solution for the YM-CG in an interval I so that $\|A\|_{S^1} \leq \epsilon$, and $B \in S^s$ a solution to the linearized equation, with $\frac{1}{2} < s \leq 1$. Then the following properties hold:*

$$\|\nabla B_0\|_{L^2 \dot{H}^{s-\frac{1}{2}}} \lesssim \|B\|_{S^s} \tag{5.18}$$

$$\|\Box A_x\|_{L^2 \dot{H}^{s-\frac{3}{2}}} \lesssim \epsilon \|B\|_{S^s} \tag{5.19}$$

Next we prove Proposition 5.4 with $b = 0$, as well as Proposition 5.4 5.5. The proof of the case $b > 0$ of Proposition 5.4 is postponed for later in this section. We remark that while the case $b = 0$ is frequently used, the stronger bound for $b > 0$ is used just once, later in the paper, in estimating the error term $E_{1,out}$ in Section 9.

Proof of Proposition 5.4 for $b = 0$. a) We begin with the A_0 bound, where we first estimate the right hand side in the equation (1.7). Using Sobolev embeddings we have the dyadic estimate with off-diagonal decay

$$\begin{aligned} \|P_k[A_{j,k_1}, \partial_0 A_{j,k_2}]\|_{L^2\dot{H}^{-\frac{1}{2}}} &\lesssim 2^{-\frac{1}{6}(k_{max}-k_{min})} \||D_x|^{\frac{1}{6}} A_{j,k_1}\|_{L^2L^6} \|\partial_0 A_{j,k_2}\|_{L^\infty L^2} \\ &\lesssim 2^{-\frac{1}{6}(k_{max}-k_{min})} \|A_{j,k_1}\|_{S^1} \|A_{j,k_2}\|_{S^1} \end{aligned} \tag{5.20}$$

After dyadic summation this gives

$$\|[A_{j,k_1} \partial_0 A_{j,k_2}]\|_{\ell^1 L^2\dot{H}^{-\frac{1}{2}}} \lesssim \|A_j\|_{S^1} \|A_j\|_{S^1} \lesssim \epsilon^2$$

Now we solve the equation (1.7) perturbatively in $\ell^1 L^2 \dot{H}^{\frac{3}{2}}$, estimating the terms $[A_j, [A_j, A_0]]$ and $[A_j, \partial_j A_0]$ in the same manner as above, appropriately using Sobolev embeddings to gain off-diagonal decay in frequency.

We need to separately prove the $\partial_t A_0$ bound, for which we use the equation (1.8). Then it suffices to prove estimates of the form

$$\begin{aligned} \|[\partial_0 A_0, A_j]\|_{\ell^1 L^2\dot{H}^{-\frac{1}{2}}} &\lesssim \|\partial_0 A_0\|_{L^2\dot{H}^{\frac{1}{2}}} \|A_j\|_{\ell^2 L^\infty \dot{H}^1} \\ \|[A_0, \partial_0 A_j]\|_{\ell^1 L^2\dot{H}^{-\frac{1}{2}}} &\lesssim \|A_0\|_{L^2\dot{H}^{\frac{3}{2}}} \|\partial_0 A_j\|_{\ell^2 L^\infty L^2} \end{aligned}$$

These are also easily proved via dyadic estimates with off-diagonal decay, which in turn are obtained using Sobolev embeddings.

b) We separately consider each of the terms on the right in the equation (1.6) for A_x. exactly as in case (a), using the bound (5.16) for the terms containing A_0. Then the $b = 0$ case of (5.17) follows exactly as in case (a). We remark that the null condition is not used at all here.

□

Proof of Proposition 5.5. a) This is similar to the proof of the previous proposition. One only needs to combine the bound (5.16) and the energy bound (2.2) for A_0 with Strichartz estimates for A_x and B_x and Sobolev embeddings in order to solve the equations (1.11) and (1.12) perturbatively in $L^2\dot{H}^{s+\frac{1}{2}}$, respectively $L^2\dot{H}^{s-\frac{1}{2}}$.

b) This is similar to the corresponding bound in the $b = 0$ case of the previous proposition. The terms on the right in (1.10) are similar to those in (1.11), so exactly the same estimates apply. □

5.3. **The perturbative bounds in Proposition 4.1, 4.4.** We primarily discuss Proposition 4.1 here, as the numerology is simpler. As the terms in G_k are similar to those in F_k, the proof of Proposition 4.4 is completely similar. However, we remark that, since we work with F_k and G_k term by term, one can view Proposition 4.1 as a special case of Proposition 4.4, for $s = 1$.

Proof of Proposition 4.1. We will successively consider all terms in F, taking into account the following observations:

(a) All estimates below are consequences of the corresponding dyadic estimates. Hence, in order to gain the control of the frequency envelope for the output F it suffices to obtain an off-diagonal gain in each of the expressions we consider.

(b) The estimates in the proposition are restricted to a time interval I. However, this does not cause any difficulties since both the Strichartz bounds and the estimate (5.1) are equally valid in I. Further, we recall that by Proposition 3.2 we can readily restrict S and N functions to time intervals.

(c) Due to the Leray projector and the identity

$$F_j = \sum_k \triangle^{-1}\partial_k\big(\partial_k F_j - \partial_j F_k\big)$$

valid for divergence free vector fields F, it suffices to estimate the curl of F_k. This observation will be used for the first term below, but not for the rest.

1. The term $[A_i, \partial_j A_i]$**.** Its curl is a null form $\mathcal{N}(A_i, A_i)$, therefore it remains to produce an N bound for the expression $|D_x|^{-1}\mathcal{N}(A_i, A_i)$. But this is a direct consequence of Proposition 5.1, with a suitable off-diagonal gain.

2. The term $[A_j, \partial_j A_i]$, high-high and and high-low interactions. Here we use the Coulomb condition $\partial_j A_j = 0$ to write

$$[A_j, \partial_j A_i] = [\partial_k(\triangle^{-1}\partial_k A_j), \partial_j A_i] = [\partial_k(\triangle^{-1}\partial_k A_j), \partial_j A_i] - [\partial_j(\triangle^{-1}\partial_k A_j), \partial_k A_i]$$

which is of the form $\mathcal{N}(|D_x|^{-1}A, A)$ where the high frequency term is hit by $|D_x|^{-1}$. Then the desired bound is again a consequence of Proposition 5.1, with off-diagonal gain.

3. The term $[\partial_0 A_0, A_i]$**.** This is a Strichartz term. Precisely, we can use $\partial_0 A_0 \in L^2\dot{H}^{\frac{1}{2}}$ as in (5.16) together with the L^2L^6 Strichartz bound for A_i and Sobolev embeddings to place it in L^1L^2, with off-diagonal gain.

4. The term $[A_0, \partial_t A_i]$, high-high and and high-low interactions. Here one uses $A_0 \in L^2\dot{H}^{-\frac{1}{2}}$ and $\nabla^{-\frac{3}{2}}\partial_t A_i \in L^2L^\infty$ to place the output into L^1L^2.

5. The commutator term $[P_k, ad(A^\alpha_{<k})]\partial_\alpha A_j$**.** This is equivalent to an expression of the form

$$2^{-k}[|D_x|A^\alpha_{<k}, \partial_\alpha A_k]$$

For $\alpha \neq 0$ this gives, as in Case 2, a null form of the type $2^{-k}\mathcal{N}(A_{<k}, A_k)$ which is handled via Proposition 5.1. For $\alpha = 0$ it is equivalent to

$$2^{-k}[\nabla A_{0,<k}, \partial_t A_k]$$

which is a Strichartz term as in Case 3. Both cases have some off-diagonal gain.

6. The cubic term $[A_j, [A_j, A_i]]$**.** This is placed in L^1L^2 via Strichartz estimates and Sobolev embeddings. The off-diagonal gain is a consequence of the fact that there is a range of Strichartz estimates which can be used in order to obtain the L^1L^2 bound. $\square$

5.4. **The proof of Proposition 5.4 for** $b > 0$**.** We consider the paradifferential decomposition of the nonlinearity in the wave equation for A_j as in (4.3). For the F component we already have the bound in Proposition 4.1, more precisely (4.7), which suffices for all $0 \leq b < \frac{1}{2}$. Hence it remains to bound the expression

$$\| \sum_{k_1 < k - C} [A_{k_1,\alpha}, \partial^\alpha A_{k_2}] \|_{\ell^1 X^{b-\frac{1}{2},-b}} \lesssim \epsilon^2 \tag{5.21}$$

We first dispense with some good portions of this expression. First, by using an $L^2 L^\infty$ bound for the first factor we obtain

$$\| [A_{k_1,\alpha}, \partial^\alpha A_{k_2}] \|_{L^2} \lesssim 2^{\frac{k_1}{2}} (\|A_{k_1,0}\|_{L^2 \dot{H}^{\frac{3}{2}}} + \|A_{k_1,x}\|_{S^1}) \|A_{k_2}\|_{S^1}$$

which has off-diagonal decay when measured in $X^{b-\frac{1}{2},-b}$ at modulations $j \geq k_1 - C$ in the output. It remains to consider low modulations in the output, namely

$$Q_{<k_1-C}[A_{k_1,\alpha}, \partial^\alpha A_{k_2}].$$

We can peel off some further part of this, using the estimate

$$\| (I - \mathcal{H}^*) Q_{<k_1-C}[A_{k_1,\alpha}, \partial^\alpha A_{k_2}] \|_N \lesssim (\|A_{k_1,0}\|_{L^2 \dot{H}^{\frac{3}{2}}} + \|A_{k_1,x}\|_{S^1}) \|A_{k_2}\|_{S^1}$$

which is a consequence of (5.3) and (5.6), and again suffices for all $b < \frac{1}{2}$. Thus we have reduced the problem to an estimate for

$$H^* Q_{<k_1-C}[A_{k_1,\alpha}, \partial^\alpha A_{k_2}] = \sum_{j<k_1} Q_{<j-C}[Q_j A_{k_1,\alpha}, Q_{<j-C} \partial^\alpha A_{k_2}]$$

For each j, this fixes the angle θ between the two factors to $\theta \approx 2^{-(k_1-j)/2}$, so we can localize to angles of this size. Note carefully that these angles will be essentially disjoint on the high frequency side, but they will be overlapping on the low frequency side.

From here on we can no longer view this as a bilinear estimate for two S^1 functions. This is not just a technical difficulty; the direct bilinear null form estimate for two S^1 functions will in effect be false for $\delta < \frac{1}{4}$, which is exactly the threshold we need to cross.

To bypass this difficulty we need to use the fact that (A_0, A_x) are not arbitrary $L^2 \dot{H}^{\frac{3}{2}}$, respectively S^1 functions, but are solution for the Yang-Mills equation. Thus we can reiterate, and use again the equation (1.6) specifically for the low frequency factor A_{k_1}. Here we can take advantage of the Z norm. We will consider A_0 and A_x separately:

a) The contribution of A_0. The analysis is simpler in this case. We simply observe that, once (5.16) is proved, we can use it to expand it to a range of mixed norm spaces as follows:

$$\| |D_x|^{\frac{3}{p}} A_0 \|_{\ell^1 L^{p'} L^p} \lesssim \epsilon^2, \qquad 2 \leq p < \infty \tag{5.22}$$

We remark that this bound fails when $p = \infty$ (precisely, we can only control the ℓ^∞ norm in that case). This is why in the study of the Yang-Mills equation we cannot simply think of A_0 as directly perturbative, and is closely related to the coupling of A_0 with A_x in the second null condition leading to the trilinear estimates in the last section of the paper.

To prove (5.22), we only discuss the inhomogeneous term in the A_0 equation, as the terms involving A_0 are similar but simpler. For this, it suffices to prove the $L^1 L^\infty$ counterpart of

(5.20) without off-diagonal decay; then by interpolation we gain the off-diagonal decay for all intermediate p's, and conclude as above. Precisely, we claim that

$$\| |D|^{-2} P_k [A_{j,k_1}, \partial_0 A_{j,k_2}] \|_{L^1 L^\infty} \lesssim \|A_{j,k_1}\|_{S^1} \|A_{j,k_2}\|_{S^1} \tag{5.23}$$

The case of unbalanced frequency interactions is easy, just by using $L^2 L^\infty$ Strichartz bounds for both factors. The more delicate case is that of *high* $\times$ *high* $\to$ *low* interactions, where $k < k_1 = k_2$. There simply using $L^2 L^\infty$ for both factors would yield a bad $2^{2(k_1-k)}$ bound. To remedy this, we partition both A_{k_1} and A_{k_2} in spatial frequency with respect to a lattice of cubes $\mathcal{C}_k$ of size 2^k, so that only opposite cubes will contribute to the output. Then by Cauchy-Schwarz we have

$$\begin{aligned} \| |D|^{-2} P_k [A_{j,k_1}, \partial_0 A_{j,k_2}] \|^2_{L^1 L^\infty} &\lesssim 2^{-4k} 2^{2k_1} \left(\sum_{\mathcal{C}_k} \|P_{\mathcal{C}_k} A_{j,k_1}\|^2_{L^2 L^\infty} \right) \left(\sum_{\mathcal{C}_k} \|P_{\mathcal{C}_k} A_{j,k_2}\|^2_{L^2 L^\infty} \right) \\ &\lesssim \|A_{j,k_1}\|^2_{S^1} \|A_{j,k_2}\|^2_{S^1} \end{aligned}$$

where we have used the next to last component of the $S^\omega_k(l)$ norm in (3.5) with $k = k_{1,2}$, $k' = k$ and $l' = 0$.

We can now use (5.22) to bound directly all *low* $\times$ *high* frequency interactions in the expression $[A_{0,k_1}, \partial_0 A_{x,k_2}]$. Indeed, by Sobolev embeddings we have

$$\| |D_x|^{-\frac{1}{p}} A_0 \|_{\ell^1 L^{p'} L^\infty} \lesssim \epsilon^2.$$

Using this we can estimate

$$\| |D_x|^{-\frac{1}{p}} [A_0, \partial_0 A_x] \|_{L^{p'} L^2} \lesssim \epsilon^2 \|\partial_0 A_x\|_{L^\infty L^2}$$

which gives the desired bound as in (5.17) with $b = \frac{1}{2} - \frac{1}{p}$ in view of the embedding

$$L^{p'} L^2 \subset X^{0, \frac{1}{p} - \frac{1}{2}}$$

Since p is arbitrarily large, we obtain the desired bound for all $0 \leq b < \frac{1}{2}$.

b) The contribution of A_x. Here we begin with the bounds (5.9) and (5.10), which allow us to split A_x into two components,

$$A_x = A_x^{good} + A_x^{bad}$$

where A_x^{good} satisfies a favorable Z bound,

$$\|A_x^{good}\|_{\ell^1 Z} \lesssim \epsilon^2$$

and A_x^{bad} is the remainder, namely

$$A_x^{bad} = \Box^{-1} |D_x|^{-1} \sum_{k < k_1 = k_2} \mathcal{H}_k N(A_{k_1}, A_{k_2})$$

We can use the $\ell^1 Z$ bound directly for A_x^{good} due to (5.5), which yields off-diagonal decay for all $\delta > 0$.

For A_x^{bad}, on the other hand, we have a favorable S^1 bound with off-diagonal decay, due to (5.1), and a Z bound without off-diagonal decay. Hence interpolating the $X_\infty^{1,\frac{1}{2}}$ component of the S^1 norm with the Z norm we obtain all intermediate bounds for A_x^{bad} with off-diagonal decay. Then we can conclude as in the A_0 case. This suffices for all $b < \frac{1}{2}$.

5.5. **Proof of Proposition 4.5, the bulk part.** Here we consider most of the proof of Proposition 4.5, modulo the more delicate trilinear part in Lemma 5.6. We extend B_j outside the interval I as free waves, and B_0 by zero. Then we seek to prove the bound in the proposition on the full real line. This allows us to consider modulation localizations. We decompose the bilinear form

$$[B_{\alpha,<k}, \partial^\alpha C_k] = (I - \mathcal{H}^*)[B_{\alpha,<k}, \partial^\alpha C_k] + \mathcal{H}^*[B_{\alpha,<k}, \partial^\alpha C_k]$$

In the first term we separate the B_j and B_0 components. For B_j we use the S norm bound, together with the null condition and the estimate (5.3). For B_0 we use the $L^2\dot{H}^{s+\frac{1}{2}}$ bound in (5.6). It remains to consider the second term, for which the B_j and B_0 terms can no longer be separated:

Lemma 5.6. *Suppose that $B \in S^s$ solves the linearized equation* (1.10) *in a time interval I. Extend B_j outside I as free waves, and B_0 by zero. Then for $s < 1$, close to* 1 *we have the global estimate*

$$\|\mathcal{H}^*[B_{\alpha,<k}, \partial^\alpha C_k]\|_{N^{s-1}} \lesssim \epsilon \|B\|_{S^s} \|C_k\|_{S^1} \tag{5.24}$$

This remaining lemma is proved in Section 10.

6. The gauge transformation

This section is devoted to the proof of Proposition 4.2.

6.1. **Equivalent formulations.** A first difficulty we encounter in the proof of the proposition is that the equations for B_j are coupled via the Leray projection. Fortunately, it turns out that the coupling is perturbative, and we can discard the projector and work with the uncoupled equations:

Proposition 6.1. *Assume that A is a solution to the Yang-Mills equation in Coulomb gauge which satisfies*

$$\|A_j\|_{S^1} \leq \epsilon \ll 1.$$

Then for the equation

$$\Box B_{j,k} + 2[A_{\alpha,<k}, \partial^\alpha B_{j,k}] = F_{j,k} \tag{6.1}$$

we have the following linear estimate:

$$\|B_k\|_{S^1} \lesssim (\|F_k\|_N + \|B_k[0]\|_{\mathcal{H}}) \tag{6.2}$$

To transition from this to Proposition 4.3 it suffices to estimate the difference, namely

$$\|\Delta^{-1}\partial_j[\partial_l A_{\alpha,<k}, \partial^\alpha B_{l,k}]\|_N \lesssim \|A_{\alpha,<k}\|_{S^1} \|B_{l,k}\|_{S^1}$$

(using the null condition via $\nabla \cdot B = 0$). This is a pure S bound as we have an extra derivative on the low frequency, and follows by (5.1)..

In view of the estimates in Proposition 4.1, the frequency localized result in Proposition 6.1 is equivalent to the following nonlocalized version:

Proposition 6.2. *Assume that A is a solution to the Yang-Mills equation in Coulomb gauge which satisfies*

$$\|A_j\|_{S^1} \leq \epsilon \ll 1.$$

Then for the equation

$$\Box_A B = F$$

we have the following linear estimate:

$$\|B\|_{S^1} \lesssim (\|F\|_N + \|B_k[0]\|_{\mathcal{H}}) \tag{6.3}$$

Further, in view of the same estimates in Proposition 4.1, the last proposition is equivalent to the existence of a good parametrix for the corresponding paradifferential problem, see the proof of Theorem 5 in [16]:

Proposition 6.3. *Assume that A is a solution to the Yang-Mills equation in Coulomb gauge which satisfies*

$$\|A_j\|_{S^1} \leq \epsilon \ll 1.$$

Then for each frequency localized initial data $(B_{0k}, B_{1k}) \in \mathcal{H}$ and inhomogeneous term $F_k \in N$ there exists an approximate solution B_k for the equation (4.8)*, in the sense that:*

(i) We have the following linear estimate:

$$\|B_k\|_{S^1} \lesssim (\|F_k\|_N + \|(B_{0k}, B_{1k})\|_{\mathcal{H}}) \tag{6.4}$$

(ii) We have the small error estimates:

$$\|B_k[0] - (B_{0k}, B_{1k})\|_{\mathcal{H}} + \|\Box B_k + 2[A_{\alpha,<k}, \partial^\alpha B_k] - F_k\|_N \lesssim \epsilon(\|F_k\|_N + \|(B_{0k}, B_{1k})\|_{\mathcal{H}}) \tag{6.5}$$

6.2. **Heuristic considerations.** Naively, our goal is to "gauge out" the magnetic potential, i.e. to find a suitable transformation, which we call *the renormalization operator*, which, up to small errors, interchanges the magnetic wave equation with the flat d'Alembertian. We now outline several considerations which eventually lead to our renormalization operators.

1. Scalar conjugations. We would like to make a gauge transformation

$$C_k = O_{<k}^{-1} B_k O_{<k}$$

where $O_{<k}$ is a $\mathbf{G}$ valued map which is also localized at lower frequency, in order to turn the above equation into

$$\Box C_k = error$$

A direct computation gives

$$\Box C_k = O_{<k}^{-1}(\Box B_k - [\partial_\alpha O_{<k} O_{<k}^{-1}, \partial^\alpha B_k] + l.o.t.)O_{<k}$$

where in "lower order terms" we have included expressions where both derivatives apply to the lower frequency term $O_{<k}$. To insure cancellation here we would need to require that

$$\partial_\alpha O_{<k} O_{<k}^{-1} = -A_{<k,\alpha} \tag{6.6}$$

Solving this exactly would require the connection A to have zero curvature, which is obviously unacceptable.

2. Pseudodifferential renormalizations. The first remedy to the above failure of complete integrability is then to allow the conjugation by O to be a pseudodifferential operator, whose symbol $O(t, x, \xi)$ would then have to satisfy

$$\partial_\alpha O_{<k} O_{<k}^{-1} \xi^\alpha \approx -A_{<k,\alpha} \xi^\alpha \tag{6.7}$$

Algebraically this means that for each ξ we renormalize A_α in a single direction, which is now possible.

However, from an analytic perspective this implies that the symbol of O will have singularities associated to space-time frequencies η so that $\eta^\alpha \xi_\alpha = 0$. To bypass this second difficulty we observe that solutions to the linear wave equation are localized in frequency on the null cone $\xi_\alpha \xi^\alpha = 0$, while the leading part of A_α are also primarily localized on the cone $\eta_\alpha \eta^\alpha = 0$. This is useful because when both ξ and η are on the cone, the expression $\eta^\alpha \xi_\alpha$ cannot vanish unless ξ and η are collinear.

To take advantage of the above observation, we first note that we are in a paradifferential situation where $|\eta| \ll |\xi|$, therefore the two cones $\xi_0 = \pm|\xi|$ are completely uncoupled, and will be renormalized separately using different parametrices $O_\pm$. In particular this will allow us to work with symbols $O_\pm(t, x, \xi)$ which do not depend on ξ_0, therefore they act separately on time slices. Thus we replace (6.7) by

$$(\omega_j \partial_j \pm \partial_0) O_{<k,\pm} O_{<k,\pm}^{-1} \approx -(\omega_j A_{<k,j} \pm A_{<k,0}), \qquad \omega = \xi' |\xi'|^{-1} \tag{6.8}$$

3. Pseudodifferential vs. nonlinear: divide and conquer. Above it was easy to replace ξ_0 by $\pm|\xi|$, but, due to the nonlinear nature of the expression on the left, it is far less straightforward to do the same for η. In order to uncouple the pseudodifferential and nonlinear aspects of the analysis, we introduce an intermediate step, namely

$$(\omega_j \partial_j \pm \partial_0) O_{<k,\pm} O_{<k,\pm}^{-1} \approx (\omega_j \partial_j \pm \partial_0) \Psi_{<k,\pm} \approx -(\omega_j A_{<k,j} \xi^j \pm A_{<k,0}) \tag{6.9}$$

The transition from A to Ψ is pseudodifferential but linear, therefore appropriately (so that only differential operators in time are used) replacing η_0 by $|\eta'|$ we can rewrite the second part of the above relation as

$$(\partial_j^2 - (\omega_j \partial_j)^2) \Psi_{<k,\pm} \approx (\partial_0 \pm p_j \omega_j)(\omega_j A_{<k,j} \xi^j \pm A_{<k,0}) \tag{6.10}$$

This transition is similar to the related step in the previous Maxwell-Klein Gordon result [16].

The step from Ψ to O, on the other hand, is more algebraic in nature, and resembles the similar step in the study of wave maps, see [37]. Precisely, for fixed ω we seek to have the more general approximate relation

$$\nabla O_{<k,\pm} O_{<k,\pm}^{-1} \approx \nabla \Psi_{<k,\pm}$$

Differentiating with respect to the frequency parameter $h < k$ we obtain

$$\nabla(\partial_h O_{<h,\pm} O_{<h,\pm}^{-1}) + [\partial_h O_{<h,\pm} O_{<h,\pm}^{-1}, \nabla O_{<h,\pm} O_{<h,\pm}^{-1}] \approx \nabla \Psi_h$$

The second term on the left is quadratic, and has the added feature that the derivative applies to the lower frequency factor. Hence it is natural to discard it. Then it is natural to obtain O by integrating Ψ_h with respect to the frequency parameter h, i.e.

$$\partial_h O_{<h} O_{<h}^{-1} = \Psi_h \tag{6.11}$$

which is a well defined $\mathbf{G}$ valued evolution.

4. Perturbative vs. renormalizable. The last question we need to address is whether we need to feed all or only part of A into the construction of the renormalization operators. For simplicity one might attempt first the former, but, as it turns out, there are two distinct obstructions for this strategy. Of course, the downside of choosing the latter is that the remaining part of A needs to be treated perturbatively.

The first issue is related to the symbol regularity for O. We observe that even with ξ and η restricted to the null cones, the expression $\eta^\alpha \xi_\alpha = 0$ can still vanish but only when ξ and η are collinear. This is the well-known difficulty of *small angle interactions*. To avoid the corresponding symbol singularities, we will excise the small angle interactions from the linear flow (4.3) and treat them perturbatively; this is where the null condition comes in handy. Unfortunately, it is too much to ask to uniformly excise the small angle interactions, and instead we do this in a frequency dependent fashion. Precisely, we will treat perturbatively only the interactions at angles

$$|\angle(\xi,\eta)| \lesssim (|\eta|/|\xi|)^\delta$$

where δ is a universal small parameter. This considerations will affect the linear step in the above construction, i.e. the transition from A to Ψ.

The second issue is related to the fact that the expression $\partial^\alpha \Psi \xi^\alpha$ vanishes in frequency on the hyperplane $\eta_\alpha \xi_\alpha = 0$. Thus, it cannot at all cancel A in the region near this hyperplane. It follows that, in order for our strategy to work, the portion of A near this hyperplane must be perturbative. But then it is pointless (and indeed counterproductive) to allow it to participate in the construction of the renormalization operator. Further, A_0's leading contribution lies in this region. Thus it is natural to place A_0 fully on the perturbative side.

6.3. **The parametrix.** Here we define the parametrix for $\Box_A$ that yields the proof of Proposition 6.3. By scaling we can assume that $k = 0$ in the Proposition, and drop it from the notations. For the rest of the section we will use $k < 0$ to denote dyadic frequencies for A, Ψ and O.

Following the above heuristics, we begin with ξ of size $O(1)$ and $\omega = \xi/|\xi|$. Then we decompose $A_{j,<0}$ into a leading part $A^{main,\pm}_{j,<0}$ and a perturbative part $A^{pert,\pm}_{j,<0}$ in a fashion which depends on ω. Here the choice of $\pm$ sign corresponds to the two cones $\tau \pm |\xi| = 0$.

The first difficulty we face is that A_j are a-priori only defined in a fixed time interval I, while our analysis uses many modulation localizations, which are nonlocal in time. To address this issue, we start with A_j in I, and extend them in time outside I as free waves. By Proposition 3.2, such an extension does not increase significantly the S^1 norm of A.

Denoting the Fourier variables for A by (σ, η), the two relevant geometric objects are the null cone $|\sigma| = |\eta|$ and the null plane $\sigma \pm \eta \cdot \omega = 0$.

It is natural to consider the two components of η, namely $\eta \cdot \omega$ and $\eta_\perp = \eta - \omega \eta \cdot \omega$. We first define a partition of the Fourier space

$$\mathbb{R}^{4+1} = D^{\omega,\pm}_{cone} \cup D^{\omega,\pm}_{null} \cup D^{\omega,\pm}_{out}$$

where the three regions are homogeneous, symmetric with respect to the origin and

$$\begin{aligned}
D^{\omega,\pm}_{cone} &= \{\mathrm{sgn}(\sigma)(\sigma \pm \eta\cdot\omega) < -\frac{1}{16}|\eta|^{-1}(|\eta_\perp|^2 + |\sigma \pm \eta\cdot\omega|^2)\} \cap \{|\sigma| < 4|\xi|\},\\
D^{\omega,\pm}_{null} &= \{|\sigma \pm \eta\cdot\omega| \lesssim \frac{1}{8}|\eta|^{-1}(|\eta_\perp|^2 + |\sigma \pm \eta\cdot\omega|^2)\},\\
D^{\omega,\pm}_{out} &= \{\mathrm{sgn}(\sigma)(\sigma \pm \eta\cdot\omega) > \frac{1}{16}|\eta|^{-1}(|\eta_\perp|^2 + |\sigma \pm \eta\cdot\omega|^2)\} \cup \{|\sigma > 2|\xi|\}
\end{aligned}$$

Correspondingly we consider a partition of unit

$$1 = \Pi^{\omega,\pm}_{cone} + \Pi^{\omega,\pm}_{null} + \Pi^{\omega,\pm}_{out}$$

where the regularity of these symbols degenerates where (σ,η) and $(\mp 1,\omega)$ are collinear,

$$\partial^\alpha_{\sigma,\eta_{perp}} \partial^\beta_{\eta_{||}} |\Pi_*^{\omega,\pm}| \lesssim \left(\frac{|\eta|}{|\eta_\perp| + (|\eta||\sigma \pm \eta\cdot\omega|)^{\frac{1}{2}}} \right)^{2|\alpha|+|\beta|}$$

Our second partition is with respect to angles. Given an angle $0 < \theta < \Pi/2$ we partition the Fourier space as

$$\mathbb{R}^{4+1} = D^{\omega,\pm}_{<2\theta} \cup D^{\omega,\pm}_{>\theta/2}$$

where

$$D^{\omega,\pm}_{<2\theta} = \{\angle(\omega, -\eta \operatorname{sgn}(\sigma)) < 2\theta\}, \qquad D^{\omega,\pm}_{<2\theta} = \{\angle(\omega, -\eta \operatorname{sgn}(\sigma)) > \theta/2\}$$

Correspondingly we define a partition of unit

$$1 = \Pi^{\omega,\pm}_{<\theta} + \Pi^{\omega,\pm}_{>\theta}$$

with the obvious symbol regularity.

Now we are ready to define the decomposition of $A_{j,<0}$, namely

$$A_{j,<0}(t,x) = A^{main,\pm}_{j,<0}(t,x,\xi) + A^{pert,\pm}_{j,<0}(t,x,\xi)$$

where

$$\begin{aligned} A^{main,\pm}_{j,<0}(t,x,\xi) =& \Pi^{\omega,\pm}_{>|\eta|^\delta} \Pi^{\omega,\pm}_{cone} A_{j,<0} \\ A^{pert,\pm}_{j,<0}(t,x,\xi) =& (\Pi^{\omega,\pm}_{<|\eta|^\delta} \Pi^{\omega,\pm}_{cone} + \Pi^{\omega,\pm}_{null} + \Pi^{\omega,\pm}_{out}) A_{j,<0} \end{aligned}$$

Here we make two observations. First, the size of the excised angle decreases with the size of the frequency $|\eta|$. This is needed in order to guarantee decay of the perturbative errors as $|\eta| \to 0$. Secondly, even though $\Pi^{\omega,\pm}_{>|\eta|^\delta}$ has a jump discontinuity at $\sigma = 0$, the symbol $\Pi^{\omega,\pm}_{cone}$ vanishes at $\sigma = 0$ so the discontinuity disappears.

Next we use the symbols $A^{main,\pm}_{j,<1}$ to define the $\mathfrak{g}$ valued zero homogeneous symbols $\Psi_\pm = \Psi_{<0,\pm}$, by

$$\Psi_\pm(t,x,\xi) \;=\; -L^\omega_\mp \Delta^{-1}_{\omega^\perp} A^{main,\pm}_{j,<1} \tag{6.12}$$

where

$$L^\omega_\pm = \partial_t \pm \omega\cdot\nabla_x, \qquad \Delta_{\omega^\perp} = \Delta - (\omega\cdot\nabla_x)^2,$$

Later in the analysis we will also use the frequency localized functions $A^{main,\pm}_{j,k}$ and $\Psi_{\pm,k}$ defined for a continuous dyadic parameter $h < 0$ so that

$$A^{main,\pm}_{j,<k} = \int_{-\infty}^k A^{main,\pm}_{j,<h} dh, \qquad \Psi_{\pm,<k} = \int_{-\infty}^k \Psi_{\pm,h} dh \tag{6.13}$$

Once we have the $\mathfrak{g}$ valued symbols $\Psi_{\pm,k}$, we define the zero homogeneous $\mathbf{G}$ valued symbols $O_{\pm,<k}(t,x,\xi)$ by solving the following differential equation on the Lie group $\mathbf{G}$,

$$\frac{d}{dk} O_{<k,\pm} O^{-1}_{<k,\pm} = \Psi_{\pm,k}, \qquad O_{-\infty,\pm} = const \tag{6.14}$$

Here the ode is solved separately for each (x,ξ), and the solution is uniquely determined up to multiplication $O \to OU$ with $U = U(x,\xi)$ an arbitrary $\mathbf{G}$-valued function. While a-priori U may depend on x and ξ, we can partially eliminate this dependence by requiring that

$$\lim_{k\to-\infty} \|\partial_x O_{<k,\pm}(t,x,\xi)\|_{L^\infty} = 0, \tag{6.15}$$

This uniquely determines $O_\pm$ up to multiplication with respect a field $U(\xi)$. We will allow this ambiguity to remain; all of our results will be invariant with respect to such a conjugation.

To construct the parametrix for the equation (4.8) we fix a large universal constant κ (e.g. $\kappa = 10$), and use the symbols

$$O_\pm(x, D) := O_{\pm,<-\kappa}(x, D)$$

and the associated operators $Op(Ad(O_\pm))(x, D)$. To do this we conjugate the constant coefficient wave flow with respect to the pair $Op(Ad(O_\pm))(x, D)$ on the left, respectively their adjoints $Op(Ad(O_\pm^{-1}))(D, y)$ on the right. The $\pm$ operators apply to the $\pm$ waves.

It is important to remark here on a minor technical point that will affect the exact definition of the parametrix. Precisely, our parametrix should take frequency one functions to frequency one functions. However, even though the symbols $\Psi_{\pm,k}$ have sharp frequency localization, the symbols $O_{\pm,<k}$ are defined in a nonlinear fashion and do not fully inherit this property. Thus, instead of using directly the operators $Op(Ad(O_\pm))(x, D)$ in our parametrix, we need to relocalize these symbols at frequencies much smaller than 1; for this we use the notation

$$(Ad(O_\pm(x, \xi)))_{<0} = P(|D_x| \ll 1)Ad(O(x, \xi)),$$

which is nothing but a localized average of $O_\pm(x, \xi)$ on the unit spatial scale. We further remark that this truncation is largely harmless, because the symbols $O_\pm$ exhibit rapid decay with favorable bounds at all frequencies much larger than $2^{-\kappa}$. This issue is discussed in detail in [16], and we will only go over it lightly in here.

The approximate solution B will have the form

$$\begin{aligned} B(t) = \sum_\pm \frac{1}{2} Op(Ad(O_\pm)_{<0})(t, x, D)e^{\pm it|D|}Op(Ad(O_\pm^{-1})_{<0})(D, 0, y)(B_0 \pm i|D|^{-1}B_1) \\ + Op(Ad(O_\pm)_{<0})(t, x, D)\frac{1}{|D|}K^\pm Op(Ad(O_\pm^{-1})_{<0})(D, s, y)F \end{aligned} \tag{6.16}$$

where

$$K^\pm f(t) = \int_0^t e^{\pm i(t-s)|D|} f(s)ds$$

represents the solution to

$$(\partial_t \mp i|D|)u = f, \qquad u(0) = 0$$

By analogy with the MKG problem, we need to prove the following bounds:

Theorem 3. *The frequency localized renormalization operators $Op(Ad(O_\pm)_{<0})(t, x, D)$ have the following mapping properties with $Z \in \{N_0, L^2, N_0^*\}$:*

$$Op(Ad(O_\pm)_{<0})(t, x, D) : \quad Z \to Z \ , \tag{6.17}$$

$$\partial_t Op(Ad(O_\pm)_{<0})(t, x, D) : \quad Z \to \epsilon Z \ , \tag{6.18}$$

$$Op(Ad(O_\pm)_{<0})(t, x, D)Op(Ad(O_\pm^{-1})_{<0})(D, y, s) - I : \quad Z \to \epsilon Z, \tag{6.19}$$

$$Op(Ad(O_\pm)_{<0})\Box - \Box^p_{A_{<0}} Op(Ad(O_\pm)_{<0}) : \quad S^\sharp_{0,\pm} \to \epsilon N_{0,\pm} \ . \tag{6.20}$$

$$Op(Ad(O_\pm)_{<0}) : \quad S^\sharp_0 \to S_0 \ , \tag{6.21}$$

where

$$\Box^p_{A_{<0}} = \Box + 2ad(A_{\alpha,<0})\partial^\alpha.$$

We remark that, as we have constructed it above, O is defined globally in time, and is based on the free wave extension of A_j outside the interval I. All the bounds in the above theorem will also be proved globally in time; indeed, with the exception of the error estimate (6.20), only the S^1 norm of A_x and the Coulomb Gauge condition are used. However, in order to prove the bound (6.20) we will need to use the Yang-Mills equation for A_x in I, as well as the definition of A_0 in terms of A_x, also in I.

The rest of the paper are devoted to the proof of the theorem. For the remainder of this section we use the Theorem to conclude the proof of Proposition 6.3:

Proof of Proposition 6.3. This is completely analogous to the proof of Theorem 4 in [16]. We define the approximate solution via (6.16). Then the bound (6.4) follows from (6.17), (6.21).

Next, we prove (6.5). For the homogeneous part of the parametrix at time $t = 0$, we have

$$\begin{aligned} B(0) - B_0 =& \frac{1}{2}\sum_{\pm} Op(Ad(O_\pm)_{<0})(0,x,D)Op(Ad(O_\pm^{-1})_{<0})(D,0,y)(B_0 \pm i|D|^{-1}B_1) - B_0 \\ =& \big[\frac{1}{2}\sum_{\pm} Op(Ad(O_\pm)_{<0})(0,x,D)Op(Ad(O_\pm^{-1})_{<0}(D,0,y)) - I\big](B_0 \pm i|D|^{-1}B_1) \end{aligned}$$

Thus the bound

$$\big\|B(0) - B_0\big\|_{\dot H^1} \lesssim \epsilon \big\|(B_0, B_1)\big\|_{\mathcal{H}}$$

is a consequence of (6.19) applied to $Z = L^2$. Further, the inequality

$$\big\|\partial_t B(0) - B_1\big\|_{L^2} \lesssim \epsilon\big(\big\|(B_0, B_1)\big\|_{\mathcal{H}} + \big\|F\big\|_N\big)$$

is a consequence of (6.18), (6.19), see the proof of Theorem 5 in [16]. Finally, for the the inhomogeneous term, we have the following

$$\begin{aligned} \Box B + 2[A_{\alpha,<0}, \partial^\alpha B] =& \sum_{\pm} \big[\Box^p_{A_{<0}} Op(Ad(O_\pm)_{<0})(t,x,D) - Op(Ad(O_\pm)_{<0})(t,x,D)\Box\big] B_\pm \\ &+ \frac{1}{2}\sum_{\pm} \big[Op(Ad(O_\pm)_{<0})(t,x,D)Op(Ad(O_\pm^{-1})_{<0})(D,t,y) - 1\big]F \\ &+ \frac{1}{2}\sum_{\pm} \pm\big[Op(Ad(O_\pm)_{<0})(t,x,D)|D|^{-1}Op(Ad(O_\pm^{-1})_{<0})(D,t,y) - |D|^{-1}\big]\partial_t F \\ &+ \sum_{\pm} Op(Ad(O_\pm)_{<0})(t,x,D)|D|^{-1}\partial_t Op(Ad(O_\pm^{-1})_{<0})(D,t,y)F \end{aligned}$$

where we set

$$B_\pm = e^{\pm it|D|}Op(Ad(O_\pm^{-1})_{<0})(D,0,y)(B_0 \pm i|D|^{-1}B_1) + |D|^{-1}K^\pm Op(Ad(O_\pm^{-1})_{<0})(D,s,y)F$$

The first term on the right is handled by combining (6.20) with (6.17), and the last three terms are controlled using (6.19) and (6.18).

$\square$

7. Decomposability and symbol bounds for Ψ and O

In this section we review the notion of disposability, which is a convenient technical tool allowing us to easily deal with issues related to symbol calculus, which would otherwise be quite technical in the context of our function spaces. Then we provide bounds for Ψ and O, first pointwise and then in disposable spaces.

This section uses only the spatial components $A_{j,<0}$ at low frequency. We assume throughout that this is divergence free, with $\|A\|_S \leq \epsilon$ and frequency envelope c_k. We fix the $\pm$ sign to $+$ and drop it from the notations.

7.1. **A review of the Decomposable Calculus.** First we discuss the notion of decomposable function spaces and estimates. This has originated in [23], [15].

A zero homogeneous symbol $c(t,x;\xi)$ is said to be in "decomposable $L^q(L^r)$" if $c = \sum_\theta c^{(\theta)}$, $\theta \in 2^{-\mathbb{N}}$, and:

$$\sum_\theta \| c^{(\theta)} \|_{D_\theta\left(L^q_t(L^r_x)\right)} \ < \ \infty \ , \tag{7.1}$$

where, adhering to the definition in [23] and with $n=4$ throughout, we put:

$$\| c^{(\theta)} \|_{D_\theta\left(L^q_t(L^r_x)\right)} \ = \ \Big\|\Big(\sum_{k=0}^{10n} \sum_\phi \sup_\omega \ \| b^\phi_\theta \ (\theta\nabla_\xi)^k \ c^{(\theta)} \|^2_{L^r_x} \Big)^{\frac{1}{2}} \Big\|_{L^q_t} \ . \tag{7.2}$$

Here $b^\phi_\theta(\xi)$ denotes a cutoff on a solid angular sector $\big|\xi|\xi|^{-1} - \phi\big| \leqslant \theta$ for a fixed $\phi \in \mathbb{S}^{n-1}$, and the sum is taken over a uniformly finitely overlapping collection. We define $\| b \|_{DL^q(L^r)}$ as the infimum over all sums (7.1). In [15] it is shown that the following Hölder type inequality holds:

$$\| \prod_{i=1}^m b_i \|_{DL^q(L^r)} \ \lesssim \ \prod_{i=1}^m \| b_i \|_{DL^{q_i}(L^{r_i})} \ , \quad (q^{-1}, r^{-1}) = \sum_i (q_i^{-1}, r_i^{-1}) \ . \tag{7.3}$$

In the sequel we only need a special case of decompositions provided in terms of these norms:

Lemma 7.1 (Decomposability Lemma)**.** *([16], Lemma 7.1) Let $A(t,x;D)$ be any pseudodifferential operator with symbol $a(t,x;\xi)$. Suppose A satisfies the fixed time bound:*

$$\sup_t \| A(t,x;D) \|_{L^2\to L^2} \ \lesssim \ 1 \ . \tag{7.4}$$

Then for any symbol $c(t,x;\xi) \in DL^q(L^r)$ one has the space-time bounds:

$$\begin{gathered} \| (ac)(t,x;D) \|_{L^{q_1}L^2 \to L^{q_2}(L^{r_2})} \lesssim \| c \|_{DL^q(L^r)} \ , \\ \frac{1}{q_1} + \frac{1}{q} = \frac{1}{q_2}, \qquad \frac{1}{2} + \frac{1}{r} = \frac{1}{r_2}, \qquad 1 \leq q_1, q_2, q, r, r_2 \leq \infty \end{gathered} \tag{7.5}$$

In the sequel it will also be useful for us to treat estimates for products of operators in a modular way. Recall that if $a(x,\xi)$ and $b(x,\xi)$ are symbols, then $a^r b^r - (ab)^r \approx i(\partial_x a \partial_\xi b)^r$. This formula is not exact, but it leads to an estimate, which is a simple variant of Lemma 7. 2 in [16]:

Lemma 7.2 (Decomposable product calculus)**.** *Let $a(x,\xi)$ and $b(x,\xi)$ be smooth symbols, and $\lambda > 0$. Then:*

(7.6)

$$\| a^r b^r - (ab)^r \|_{L^r(L^2)\to L^q(L^2)} \lesssim \sup_{1\le|\alpha|<N} \lambda^{|\alpha|} \| (\nabla_x a)^r \|_{L^r(L^2)\to L^{p_1}(L^2)} \sup_{1\le|\alpha|<N} \lambda^{|\alpha|} \| \nabla_\xi^\alpha b \|_{L^{p_1}L^2\to L^q L^2}$$

7.2. **Bounds for A.** Here we state the decomposability bounds for A, see [16], Lemma 7.3:

Lemma 7.3. *The functions $A_x\cdot\omega$, A_0 satisfy the following decomposability bounds:*

$$\|P_k(A_x^{(\theta)}\cdot\omega, A_0)\|_{DL^pL^\infty} \lesssim 2^{(1-\frac{1}{p})k}\theta^{\frac{5}{2}-\frac{2}{p}} \tag{7.7}$$

7.3. **Bounds for Ψ.** For the purpose of our first step we use the frame determined by $\omega = \xi|\xi|^{-1}$ and its orthogonal complement $\omega^\perp$ to describe the regularity of Ψ. We have

Lemma 7.4. *The functions $\Psi_k(t,x,\xi)$ satisfy the following bounds for fixed t and ξ:*

$$\|\nabla_{\omega^\perp}\nabla\Psi_k\|_{L^2} \lesssim c_k \tag{7.8}$$

$$\|\nabla^2\Psi_k\|_{L^2} \lesssim 2^{-\delta k}c_k \tag{7.9}$$

We also get the bounds

$$\|\nabla_\xi^N\nabla^2\Psi_k\|_{L^2} \lesssim 2^{-(N+1)\delta k}c_k. \tag{7.10}$$

We remark that, as a consequence of Bernstein's inequality, the bound (7.8) implies the pointwise bounds

$$\|\Psi_k\|_{L^\infty} \lesssim c_k \tag{7.11}$$

Also we consider L^p norms at fixed time. Fixing ξ we use the orthonormal frame associated to ξ, and the mixed norms $L^2_\omega L^6$ and $L^\infty_\omega L^3$. By Bernstein's inequality, from (7.8) we obtain

$$\|\nabla_x\Psi_k\|_{L^2_\omega L^6} + \|\nabla_\perp\Psi_k\|_{L^\infty_\omega L^3} \lesssim c_k \tag{7.12}$$

Proof. We first note that simply using the L^2 fixed time bound for ∇A does not suffice due to the presence of two inverse derivatives in (6.12). We will use the Coulomb gauge condition to cancel one of these two derivatives. Precisely, using the Coulomb gauge condition to write

$$A_{j,k}\omega_j = (\omega_j - |\Delta|^{-1}\nabla\otimes\nabla)A_{j,k} = \Delta^{-1}\nabla\nabla_{\omega^\perp}A_{j,k}$$

which is exactly what we need.

□

Next, we consider a number of decomposable estimates for the phase $\Psi(t,x;\xi)$ used to define our microlocal gauge transformations:

Lemma 7.5 (Decomposable estimates for Ψ)**.** *Let the phase $\Psi(t,x;\xi)$ be defined as in (6.12), and its angular components $\Psi^{(\theta)} = \Pi_\theta^\omega\psi(t,x;\xi)$, where $\omega = |\xi|^{-1}\xi$. Then for $q\ge 2$ and $2/q+3/r\le\frac{3}{2}$ one has:*

$$\| (\Psi_k^{(\theta)}, 2^{-k}\nabla_{t,x}\Psi_k^{(\theta)}) \|_{DL^q(L^r)} \lesssim 2^{-(\frac{1}{q}+\frac{4}{r})k}\theta^{\frac{1}{2}-\frac{2}{q}-\frac{3}{r}}\epsilon, \tag{7.13}$$

In addition, suppose that $\theta\lesssim 2^j\lesssim 1$. Then for $q,r\ge 2$ we also have

$$\| Q_{k+2j}(\Psi_k^{(\theta)}, 2^{-k}\nabla_{t,x}\Psi_k^{(\theta)}) \|_{DL^q(L^r)} \lesssim 2^{-(\frac{1}{q}+\frac{4}{r})k}2^{-(1-\frac{2}{q})j}\theta^{\frac{1}{2}-\frac{3}{r}}\epsilon, \tag{7.14}$$

Further,

$$\| \Box \Psi_k^{(\theta)} \|_{DL^2(L^\infty)} \lesssim \theta^{\frac{3}{2}} 2^{\frac{3}{2}k}, \tag{7.15}$$

In particular

$$\| (\Psi_k, 2^{-k}\nabla_{t,x}\Psi_k) \|_{DL^q(L^\infty)} \lesssim 2^{-\frac{1}{q}k}\epsilon , \qquad q > 4 , \tag{7.16}$$

$$\| Q_{k+2j}(\Psi_k, 2^{-k}\nabla_{t,x}\Psi_k) \|_{DL^q(L^\infty)} \lesssim 2^{-\frac{1}{q}k} 2^{(\frac{1}{2}-\frac{2}{q})j}\epsilon , \qquad 2 \leq q < 4 , \tag{7.17}$$

$$\| \nabla_{t,x}\Psi_k \|_{DL^2(L^r)} \lesssim 2^{(\frac{1}{2}-\frac{4}{r}-\delta(\frac{1}{2}+\frac{3}{r}))k}\epsilon , \qquad r \geq 6 , \tag{7.18}$$

Proof. Notice that the last three estimates follow from the first by summing over dyadic $2^{-\delta k} \leq \theta \lesssim 1$. For the first two bounds we interchange the t integration and the ω summation to obtain:

$$\begin{aligned} \| (\psi_k^{(\theta)}, 2^{-k}\nabla_{t,x}\psi_k^{(\theta)}) \|_{DL^q(L^r)} &\lesssim \theta^{-2} 2^{-k} \Big(\sum_\omega \| \Pi_\theta^\omega(D) A \cdot \omega \|_{L^q(L^r)}^2 \Big)^{\frac{1}{2}} \\ &\lesssim \theta^{-1} 2^{-k} \Big(\sum_\omega \| \Pi_\theta^\omega(D) A \|_{L^q(L^r)}^2 \Big)^{\frac{1}{2}}, \end{aligned}$$

where at the second step we have used the Coulomb gauge to gain another factor of θ.

Now we conclude the proof of (7.16) using the Strichartz estimate component of the S_k norms. In four space dimensions the Strichartz sharp range is given by $\frac{2}{q} + \frac{3}{r_0} = \frac{3}{2}$. Moreover, on an angular sector of size θ Bernstein's inequality gives the embedding $\Pi_\theta^\omega(D) P_k L^{r_0} \subseteq \theta^{3(\frac{1}{r_0}-\frac{1}{r})} 2^{4(\frac{1}{r_0}-\frac{1}{r})k} L^r$. Thus:

$$\Big(\sum_\omega \| \Pi_\theta^\omega(D) A_k \|_{L^q(L^r)}^2 \Big)^{\frac{1}{2}} \lesssim \theta^{\frac{3}{2}-\frac{2}{q}-\frac{3}{r}} 2^{(1-\frac{1}{q}-\frac{4}{r})k} \| A_k \|_{S_k} ,$$

and (7.16) follows.

The argument for (7.17) is simpler. The case $q = r = 2$ is immediate using the L^2 bound coming from the $X_\infty^{1,\frac{1}{2}}$ component of the S^1 norm, and the transition to larger q, r is done using Bernstein's inequality. $\square$

We wrap this section up by proving some additional symbol type bounds for the phases Ψ. These involve the variation over the physical space variables:

Lemma 7.6 (Additional symbol bounds for Ψ)**.** *Let ψ be as above. Then one has:*

$$|\Psi_{<k}(t,x;\xi) - \Psi_{<k}(s,y;\xi)| \lesssim \epsilon \log(1 + 2^k(|t-s| + |x-y|)), \tag{7.19}$$

$$|\Psi(t,x;\xi) - \Psi(s,y;\xi)| \lesssim \epsilon \log(1 + |t-s| + |x-y|) \tag{7.20}$$

$$|\partial_\xi^\alpha(\Psi(t,x;\xi) - \Psi(s,y;\xi))| \lesssim \epsilon \langle (t-s, x-y) \rangle^{|\alpha - \frac{1}{2}|\sigma}, 1 \leqslant \alpha \leqslant \sigma^{-1}. \tag{7.21}$$

Proof. We decompose as before

$$\psi_{<k}(t,x;\xi) = \sum_{j<k} \sum_{\theta > 2^{\sigma j}} \psi_j^{(\theta)}(t,x,\xi)$$

For each fixed θ and j we have by the definition of ψ and the Coulomb gauge condition

$$|\psi_j^{(\theta)}(t,x,\xi)| \lesssim \theta^{-1} 2^{-j} \sup_\omega \|\Pi_\theta^\omega A_j\|_{L^\infty}$$

Then by energy estimates for A and Bernstein's inequality we obtain

$$|\psi_j^{(\theta)}(t,x,\xi)| \lesssim \theta^{\frac{1}{2}} \|A_j[0]\|_{H^1\times L^2}, \qquad |\psi_j(t,x,\xi)| \lesssim \|A_j[0]\|_{H^1\times L^2} \tag{7.22}$$

A similar argument leads to

$$|\partial_{t,x}\psi_j^{(\theta)}(t,x,\xi)| \lesssim 2^j \theta^{\frac{1}{2}} \|A_j[0]\|_{H^1\times L^2}, \qquad |\partial_{t,x}\psi_j(t,x,\xi)| \lesssim 2^j \|A_j[0]\|_{H^1\times L^2} \tag{7.23}$$

Differentiating with respect to ξ yields θ^{-1} factors,

$$|\partial_\xi^\alpha \psi_j^{(\theta)}(t,x,\xi)| \lesssim \theta^{\frac{1}{2}-|\alpha|} \|A_j[0]\|_{H^1\times L^2}, \qquad |\partial_{x,t}\partial_\xi^\alpha \psi_j^{(\theta)}(t,x,\xi)| \lesssim 2^j \theta^{\frac{1}{2}-|\alpha|} \|A_j[0]\|_{H^1\times L^2}.$$

For the bound (7.19) we use both (7.22) and (7.23) to write for $j \leq k$

$$|\psi_{<k}(t,x;\xi) - \psi_{<k}(s,y;\xi)| \lesssim 2^j(|t-s| + |x-y|) + |k-j|$$

and then optimize the choice of j.

The proof of (7.21) is similar.

□

7.4. **Fixed time bounds for O.** Here we transfer the above bounds from Ψ to O. Precisely, we have the following

Lemma 7.7. *The following estimates hold for O, where $\perp$ below refers to derivatives in the plane $\omega^\perp$:*

$$\|P_{k'} O_{;\perp}\|_{L^2} \lesssim 2^{-k'} c_{k'} 2^{-N(k'-k)_+} \tag{7.24}$$

$$\|P_{k'} O_{;x,t}\|_{L^2} \lesssim 2^{(-\delta-1)k'} c_{k'} 2^{-N(k'-k)_+} \tag{7.25}$$

Estimates with one derivative less hold for $\partial_k O_{;x,t}$.

Proof. We treat the case of spatial derivatives, time derivatives being handled similarly. Our strategy will be to use integration in h and reiteration in the commutation relation

$$\frac{d}{dh} O_{<h;x} = \Psi_{h,x} + [\Psi_h, O_{<h;x}] \tag{7.26}$$

as well as differentiated forms of it, in order to build up successively stronger bounds for the derivatives of $O_{<h}$. In this section, mixed Lebesgue spaces $L^p L^q$ refer to the coordinates $\omega, \omega^\perp$ for the x-plane.

1. L^∞ bounds. A-priori we have

$$\|\Psi_k\|_{L^\infty} \lesssim c_k.$$

Then integration from $-\infty$ with respect to h in (7.26) gives

$$\|O_{<k;x}\|_{L^\infty} \lesssim 2^k c_k$$

Repeated differentiation similarly leads to a better high frequency bound

$$\|\partial_x^{m-1} O_{<k;x}\|_{L^\infty} \lesssim 2^{mk} c_k$$

2. $L^2 L^{12}$ bounds. Here we start with

$$\|\Psi_k\|_{L^2 L^{12}} \lesssim 2^{-\frac{3k}{4}} c_k$$

The same argument as above using (7.26) leads to

$$\|O_{<k;x}\|_{L^2L^{12}} \lesssim 2^{\frac{k}{4}} c_k \qquad \|\partial_x^{m-1} O_{<k;x}\|_{L^2L^{12}} \lesssim 2^{(m-\frac{3}{4})k} c_k$$

3. $L^\infty L^6$ bounds. Here we start with

$$\|\partial_\perp \Psi_k\|_{L^\infty L^6} \lesssim 2^{\frac{1}{2}k} c_k$$

As above, using (7.26) but only for $\partial_\perp$ derivatives we obtain

$$\|O_{<k;\perp}\|_{L^\infty L^6} \lesssim 2^{\frac{k}{2}} c_k \qquad \|\partial_x^{m-1} O_{<k;\perp}\|_{L^\infty L^6} \lesssim 2^{(m-\frac{1}{2})k} c_k$$

4. L^2L^6 bounds. For this we use the bound

$$\|\Psi_k\|_{L^2L^6} \lesssim 2^{-k} c_k$$

We apply a Littlewood-Paley projector $P_{k'}$ in (7.26) and integrate in h,

$$\|P_{k'} O_{<k;x}\|_{L^2L^6} \lesssim \int_{-\infty}^{k} \|P_{k'} \Psi_{h,x}\|_{L^2L^6} + \|P_{k'}[\Psi_h, O_{<h;x}]\|_{L^2L^6}\, dh$$

The first term on the right contributes only when $h = k + O(1)$. Thus we consider two scenarios. If $k < k'$ then we combine directly the high frequency L^∞ bound for $O_{;x}$ with the L^2L^6 bound for $\Psi_{h,x}$ to obtain the rapid decay

$$\|P_{k'} O_{<k;x}\|_{L^2L^6} \lesssim c_{k'} 2^{-N(k'-k)}, \qquad k < k'$$

If $k \geq k'$ then we retain the contribution of the first term when $h = k' + O(1)$, and in addition we bound the second term for larger $h > k'$ using Bernstein's inequality as follows:

$$\|P_{k'}[\Psi_h, O_{<h;x}]\|_{L^2L^6} \lesssim 2^{\frac{3}{4}k'} \|\Psi_h\|_{L^2L^6} \|O_{<h;x}]\|_{L^2L^{12}} \lesssim c_h^2 2^{\frac{3}{4}(k'-h)}$$

Taking advantage of the decay in h, we obtain the desired bound

$$\|P_{k'} O_{<k;x}\|_{L^2L^6} \lesssim c_{k'}, \qquad k \geq k'.$$

5. $L^\infty L^3$ bounds. For this we use the bound

$$\|\partial_\perp \Psi_k\|_{L^\infty L^3} \lesssim 2^{-k} c_k$$

and argue as in the L^2L^6 case. The only difference arises in the treatment of the bilinear term for $h \geq k'$, namely

$$\|P_{k'}[\Psi_h, O_{<h;\perp}]\|_{L^\infty L^3} \lesssim 2^{\frac{1}{2}k'} \|\Psi_h\|_{L^2L^6} \|O_{<h;\perp}]\|_{L^\infty L^6} \lesssim c_h^2 2^{\frac{1}{2}(k'-h)}$$

We obtain

$$\|P_{k'} O_{<k;\perp}\|_{L^\infty L^3} \lesssim c_{k'} 2^{-N(k'-k)_+}$$

6. L^2 bounds. In this final step we use the equation

$$\frac{d}{dh} P_{k'} O_{<h;\perp} = P_{k'} \Psi_{h,\perp} + P_{k'}[\Psi_h, O_{;\perp}]$$

take L^2 norms and integrate with respect to h. For $h < k'$ the first term on the right vanishes, while for the second we have

$$\|P_{k'}[\Psi_h, O_{<h;\perp}]\|_{L^2} \lesssim \|\Psi_h\|_{L^2L^6}\|P_{k'}O_{;\perp}\|_{L^\infty L^3} \lesssim c_h^2 2^{-h}2^{-N(k'-h)}$$

Integrating we obtain

$$\|P_{k'}O_{<k;\perp}\|_{L^2} \lesssim c_{k'}2^{-k'}2^{-N(k'-k)}, \qquad k < k'.$$

It remains to consider the case $k > k'$. The first term $P_{k'}\Psi_{h,\perp}$ is nonzero only if $h = k' + O(1)$, in which case it is easily estimated using (7.8). For the second term, on the other hand, we have

$$\|P_{k'}[\Psi_h, O_{<h;\perp}]\|_{L^2} \lesssim \|\Psi_h\|_{L^2L^6}\|O_{;\perp}\|_{L^\infty L^3} \lesssim c_h^2 2^{-h}$$

which is easily integrated for $h > k'$. Thus the proof of (7.24) is complete.

7. Proof of the bound (7.25). This proof is largely similar, so we outline the change. In fact, in Step 5, a $2^{-\frac{1}{2}\delta k}$ loss in the $\partial_x\psi$ bound generates a similar loss for $O_{<k;x}$. The same loss propagates directly to Step 6.

7.5. **Fixed time bounds for** $O_{;\xi}$**.** Differentiating the functions Ψ_k with respect to ξ looses a factor of $\angle(\xi,\eta)$. Due to the angular separation, this factor is at most $2^{\delta k}$. Thus, the bounds for $O_{<k;\xi}$ are similarly related to the bounds for $O_{<k}$:

Lemma 7.8. *We have the pointwise bounds*

$$|\partial_\xi^n O_{<k;x}| \lesssim 2^{k(1-n\delta)},\ n\delta < 1, \tag{7.27}$$

as well as the L^2 bounds

$$\|P_{k'}\partial_\xi^n O_{<k;x}\|_{L^2} \lesssim 2^{k(-1-n\delta)}2^{-N(k'-k)_+} \tag{7.28}$$

The evolution equation for $O_{;\xi} = O_\xi O^{-1}$ is

$$\frac{d}{dh}O_{<h;\xi} = \Psi_{h,\xi} + [\Psi_h, O_{<h;\xi}]$$

We have a similar relation for O_x,

$$\frac{d}{dh}O_{<h;x} = \Psi_{h,x} + [\Psi_h, O_{<h;x}]$$

Differentiating the latter with respect to ξ yields

$$\frac{d}{dh}\partial_\xi O_{<h;x} = \partial_\xi\Psi_{h,x} + [\partial_\xi\Psi_h, O_{<h;x}] + [\Psi_h, \partial_\xi O_{<h;x}]$$

Since

$$|\Psi_{h,x}| + |O_{<h;x}| \lesssim 2^h, \qquad |\partial_\xi\Psi_{h,x}| \lesssim 2^{h(1-\delta)}$$

we can integrate to obtain

$$|\partial_\xi O_{<h;x}| \lesssim 2^{h(1-\delta)}$$

We further have L^2 bounds

$$\|\partial_\xi\Psi_{h,x}\|_{L^2} \lesssim 2^{h(-1-2\delta)}, \qquad \|\partial_\xi\Psi_h\|_{L^2} \lesssim 2^{h(-2-2\delta)}, \qquad \|\Psi_h\|_{L^2} \lesssim 2^{h(-2-\delta)}$$

with extra gain for further x derivatives. We can transfer these bounds to $\partial_\xi O_{<h;x}$ by using Littlewood-Paley projectors in x in the above evolution, to obtain (7.28).

We also have the commutation relation

$$\partial_\xi O_{<h;x} - \partial_x O_{<h;\xi} = [O_{<h;x}, O_{<h;\xi}] \tag{7.29}$$

Up to this point O is only uniquely determined up to a ξ dependent conjugation,

$$O_{<h}(x,\xi) \to O_{<h}(x,\xi)P(\xi)$$

At the level of $O_{<h;\xi}$ this translates to the gauge freedom

$$O_{<h;\xi}(x,\xi) \to O_{<h;\xi}(x,\xi) + O_{<h}(x,\xi)P_\xi P^{-1} O_{<h}^{-1}(x,\xi)$$

Fixing a choice of P is not necessary, as all estimates we need are invariant under such a change.

7.6. **Decomposable bounds for $O_{;x}$, $O_{;t}$.** Our goal here is to transfer decomposability bounds from Ψ to O. Precisely, we have

Lemma 7.9. *We have the following estimates:*

$$\| O_{<k;x}, O_{<k;t} \|_{DL^q(L^\infty)} \lesssim 2^{(1-\frac{1}{q})k}\epsilon \ , \qquad q > 4 \ , \tag{7.30}$$

$$\| O_{<k;x}, O_{<k;t} \|_{DL^2(L^\infty)} \lesssim 2^{\frac{1}{2}(1-\delta)k}\epsilon \ , \tag{7.31}$$

Proof. We prove the bounds for $O_{<k;x}$; those for $O_{<k;t}$ are identical. We use the evolution for $O_{k;x}$, namely

$$\partial_k O_{<k;x} = \Psi_{k,x} + [\Psi_k, O_{<k;x}]$$

We proceed in several stages:

Step 1:(*A weaker $DL^\infty L^\infty$ bound*) Using the pointwise bounds on $O_{;x}$ and its ξ derivatives, we directly conclude that (for $\delta > 0$ small enough)

$$\|O_{<k;x}\|_{DL^\infty L^\infty} \lesssim 2^{(1-n\delta)k}, \ n = 40.$$

Step 2:(*The full $DL^\infty L^\infty$ bound*) Using the above evolution we obtain the integral bound

$$\|O_{<k;x}\|_{DL^\infty L^\infty} \lesssim \|O_{<l;x}\|_{DL^\infty L^\infty} + \int_l^k \|\Psi_{h,x}\|_{DL^\infty L^\infty} + \|\Psi_h\|_{DL^\infty L^\infty}\|O_{<h;x}\|_{DL^\infty L^\infty} dh$$

By Gronwall's inequality this gives

$$\|O_{<k;x}\|_{DL^\infty L^\infty} \lesssim \|O_{<l;x}\|_{DL^\infty L^\infty} e^{\int_l^k c_h dh} + \int_l^k 2^h c_h e^{\int_h^k c_{h_1} dh_1} dh \tag{7.32}$$

But by Cauchy-Schwarz we have

$$\int_l^k c_h dh \lesssim |k-l|^{\frac{1}{2}}$$

Thus, using the weaker $DL^\infty L^\infty$ bound, the first term in (7.32) decays to zero as $l \to -\infty$. On the other hand, the leading contribution in the second term in (7.32) comes from $h = k - O(1)$. Hence we obtain the desired bound.

$$\|O_{<k;x}\|_{DL^\infty L^\infty} \lesssim 2^k c_k$$

Step 3:(*The DL^qL^∞ bound*) Using again the above evolution and the fact that, by construction, $\lim_{k\to 0} O_{<k;x} = 0$ we write

$$O_{<k;x} = \int_{-\infty}^{k} \Psi_{h,x} + [\Psi_h, O_{<h;x}]dh$$

Then we combine the DL^qL^∞ decomposability bound (7.16) for Ψ_h with the previously established $DL^\infty L^\infty$ bound for $O_{<h;x}$.

Step 4:(*The DL^2L^∞ bound*) We proceed as in the previous step, but using the bound (7.16) for Ψ_h instead.

□

7.7. **Difference bounds for** O**.** Here we seek to compare $O_{<k}(t,x,\xi)$ with $O_{<k}(s,y,\xi)$. Since both are elements of the Lie group $\mathbf{G}$, it is natural (and most useful in the sequel) to look at the product $O_{<k}(t,x,\xi)O^{-1}_{<k}(s,y,\xi)$. We have

Lemma 7.10 (Difference bounds for O)**.** *Let O be as above. Then one has:*

$$d(O_{<k}(t,x,\xi)O^{-1}_{<k}(s,y,\xi), Id) \lesssim \epsilon \log(1 + 2^k(|t-s| + |x-y|)), \tag{7.33}$$

$$d(O(t,x,\xi)O^{-1}(s,y,\xi), Id) \lesssim \epsilon \log(1 + |t-s| + |x-y|) \tag{7.34}$$

$$|\partial_\xi^n (O(t,x,\xi)O^{-1}(s,y,\xi))_{;\xi}| \lesssim \langle (t-s, x-y)\rangle^{n\delta}. \tag{7.35}$$

Proof. For the first two bounds we use the $Ad(O^{-1}(t,x,\xi))$ to interchange the order and estimate instead the distance $d(O^{-1}_{<k}(s,y,\xi)O_{<k}(t,x,\xi), Id)$. This vanishes as $k \to -\infty$, therefore we can write

$$d(O^{-1}_{<k}(s,y,\xi)O_{<k}(t,x,\xi), Id) \lesssim \int_{-\infty}^{k} |(O^{-1}_{<h}(s,y,\xi)O_{<h}(t,x,\xi))_{;h}| dh$$

But we have

$$(O^{-1}_{<h}(s,y,\xi)O_{<h}(t,x,\xi))_{;h} = O^{-1}_{<h}(t,x,\xi)(\Psi_h(t,x,\xi) - \Psi_h(s,y,\xi))O_{<h}(s,y,\xi)$$

so we obtain

$$d(O^{-1}_{<k}(s,y,\xi)O_{<k}(t,x,\xi), Id) \lesssim \int_{-\infty}^{k} |\Psi_h(t,x,\xi) - \Psi_h(s,y,\xi)| dh$$

For Ψ_h we have the bound

$$|\Psi_h(t,x,\xi) - \Psi_h(s,y,\xi)| \lesssim \epsilon \min\{1, 2^h(|x-y| + |t-s|)\}$$

Thus the bounds (7.33) and (7.34) follow after dyadic integration with respect to h.

For the third bound (7.35) we denote $V_{<k} = O^{-1}_{<k}(t,x,\xi)O_{<k}(s,y,\xi)$, and proceed in two steps. For the first step we fix k, and show that

$$|\partial_\xi^n V_{<k;\xi}| \lesssim (|t-s| + |x-y|)2^{k(1-n\delta)}, \qquad 2^k|x-y| \lesssim 1, \quad n \geq 0. \tag{7.36}$$

This bound is favorable provided that k is small enough. In the second step, we extent the range of k for which (7.35) holds by evaluating the k derivative of $\partial_\xi^n V_{<k;\xi}$.

We now proceed with the first step, where we will crucially use the bound

$$|\partial_\xi^n O_{<k;x}(t,x,\xi)| \lesssim 2^{k(1-\delta(\frac{1}{2}+n))}, \tag{7.37}$$

see (7.28). The expression $V_{<k;\xi}$ vanishes if $x = y, t = s$ so it suffices to estimate its x, t derivatives; below we do so for the x-derivatives, with similar estimates applying to the t-derivatives:

$$\partial_x \partial_\xi^n V_{<k;\xi} = \partial_\xi^{n+1} V_{<k;x} + \partial_\xi^n [V_{<k;x}, V_{<k;\xi}]$$

for which we use $V_{<k;x} = O_{<k;x}(t, x, \xi)$ to rewrite it as

$$\partial_y \partial_\xi^n V_{<k;\xi} - [V_{<k;y}, \partial_\xi^n V_{<k;\xi}] = \partial_\xi^{n+1} O_{<k;x} + \sum_{j=1}^{n} [\partial_\xi^j O_{<k;x}, \partial_\xi^{n-j} V_{<k;\xi}].$$

The last term is absent if $n = 0$, so the bound (7.36) follows directly from (7.37) by integration. Finally we close by induction integrating over x, estimating

$$|[\partial_\xi^j O_{<k;x}, \partial_\xi^{n-j} V_{<k;\xi}]| \lesssim 2^{k(1+\delta(\frac{1}{2}-j))} |x-y| 2^{k(1-\delta(\frac{1}{2}+(n-j)))} = 2^k |x-y| 2^{k(1-\delta n)}$$

So far the bound (7.35) is established in the range $2^k |x-y| \lesssim 1$. To extend it we forget about the distance between x and y and integrate instead with respect to l (the new k). First write

$$V_{<l} = W_{<l}(x) V_{<k} W_{<l}^{-1}(y)$$

where

$$W_{<l} = O_{<l}^{-1} O_{<k}$$

We have

$$V_{<l;\xi} = -W_{<l}(x) V_{<k}^{-1} W_{<l}^{-1}(y) W_{<l;\xi}(y) W_{<l}(y) V_{<k} W_{<l}^{-1}(x) + W_{<l}(x) V_{<k;\xi} W_{<l}^{-1}(y) + W_{<l;\xi}(x)$$

so repeated differentiation shows that it suffices to bound

$$|\partial_\xi^n W_{<l;\xi}(x)| \lesssim 2^{-k\delta(\frac{1}{2}+n)}, \qquad l > k, \quad n \geq 0 \tag{7.38}$$

For this we follow the previous strategy, writing

$$\partial_l \partial_\xi^n W_{<l;\xi} = \partial_\xi^{n+1} O_{<l;l} + \partial_\xi^n [O_{<l;l}, W_{<l;\xi}]$$

which leads to

$$\partial_l \partial_\xi^n W_{<l;\xi} - [\Psi_l, \partial_\xi^n W_{<l;\xi}] = \partial_\xi^{n+1} \Psi_l + \sum_{j=1}^{n} [\partial_\xi^j \Psi_l, \partial_\xi^{n-j} W_{<l;\xi}]$$

Using the bounds for Ψ we can inductively close (7.38).

□

8. L^2 bounds for the parametrix

In this section we establish a number of L^2 bounds for the renormalization operators and the parametrix. In the last part we prove the bounds (6.17), (6.18), (6.19) and (6.21). Throughout the section we assume that A is a Yang-Mills wave with $\|A\|_S \ll 1$ and frequency envelope c_k. We fix the $\pm$ sign to $+$ and drop it from the notations. Also, we shall consider unit frequencies, and put O instead of $O_{<0}$. We split the argument across several subsections.

8.1. **Oscillatory integral estimates.** We first observe that on one hand our parametrix involves operators of the form

$$T^a = Op(Ad(O^{\pm}))(t,x,D)e^{\pm i(t-s)|D|}a(|D|)Op(Ad(O^{\pm}))(D,s,y)$$

where a is localized at frequency 1. On the other hand, arguing in TT^* fashion in order to prove various L^2 estimates involving the operators $Op(Ad(O(t,x,D))$ and $Op(Ad(O_{<0}(t,x,D)^*))$, we need to consider bounds for similar operators in the special case when $t = s$.

The kernel of the operator T_a is given by the oscillatory integral

$$K^a F(t,x) = \int a(\xi)e^{\pm i(t-s)|\xi|}e^{i\xi(x-y)}(O(t,x,\xi)O^{-1}(s,y,\xi))F(s,y)(O(t,x,\xi)O^{-1}(s,y,\xi))^{-1}d\xi$$

Our main estimates for such kernels are as follows:

Proposition 8.1. *a) Assume that a is a smooth bump on the unit scale. Then the kernel K_a satisfies*

$$|K_a(t,x;s,y)| \lesssim \langle t-s\rangle^{-\frac{3}{2}}\langle |t-s| - |x-y|\rangle^{-N} \tag{8.1}$$

b) Let $a = a_C$ be a bump function on a rectangular region C of size $2^k \times (2^{k+l})^3$ with $k \leq l \leq 0$. Then

$$|K_a(t,x;s,y)| \lesssim 2^{4k+3l}\langle 2^{2(k+l)}(t-s)\rangle^{-\frac{3}{2}}\langle 2^k(|t-s| - |x-y|)\rangle^{-N} \tag{8.2}$$

If in addition $x-y$ and C have a 2^{k+l} angular separation then

$$|K_a(t,x;s,y)| \lesssim 2^{4k+3l}\langle 2^{2(k+l)}|t-s|\rangle^{-N}\langle 2^k(|t-s| - |x-y|)\rangle^{-N} \tag{8.3}$$

Proof. a) Away from a conic neighborhood of the cone $\{|t-s| = \pm|x-y|\}$ the phase

$$\Psi = \pm(t-s)|\xi| + \xi(x-y)$$

is nondegenerate. Hence applying the symbol bounds (7.21) repeated integration by parts with respect to ξ yields

$$|K^a(t,x,s,y)| \lesssim \langle (t,x)-(s,y)\rangle^{-N}, \qquad N \sim \sigma^{-1}$$

Near the cone we need to be more careful. Denoting $T = |t-s|+|x-y|$ and $R = |t-s|-|x-y|$, in suitable (polar) coordinates the operator K^a takes the form

$$K^a F(t,x) = \int (O^{-1}(t,x,\xi')O(s,y,\xi'))F(s,y)(O^{-1}(t,x,\xi')O(s,y,\xi'))^{-1}e^{iR\xi_1}e^{iT\xi'^2}\tilde{a}(\xi)d\xi$$

In ξ_1 (the former radial variable) this is a straight Fourier transform, so we get rapid decay in R. Given the bound (7.21), we can use stationary phase in ξ'. While the ξ derivatives of the $O^{-1}(t,x,\xi')O(s,y,\xi')$ part of the phase are not bounded, they only bring factors of T^σ, which is small enough not to affect the stationary phase (this works up to $\sigma = \frac{1}{2}$). We obtain

$$|K^a(t,x,s,y)| \lesssim T^{-\frac{3}{2}}(1+R)^{-N}$$

b) Away from the cone the estimate follows easily as above since the phase is nondegenerate. Near the cone we use again polar coordinates to express our oscillatory integral as above,

$$K^C F(t,x) = \int (O^{-1}(t,x,\xi')O(s,y,\xi'))F(s,y)(O^{-1}(t,x,\xi')O(s,y,\xi'))^{-1}e^{iR\xi_1}e^{iT\xi'^2}\tilde{a}_C(\xi)d\xi$$

where a_C is a bump function in a rectangle on the 2^k scale in the radial variable ξ_1 and on the 2^{k+l} scale in the angular variable ξ'. Then we can separate variables in (ξ_1, ξ'). We note that this rectangle need not be centered at $\xi' = 0$, though this is the worst case. In ξ_1 this is again a Fourier transform, so we get the factor

$$2^k \langle 2^k R\rangle^{-N}$$

In ξ' we can use stationary phase to get the factor

$$2^{3(k+l)}\langle 2^{2(k+l)T}\rangle^{-\frac{3}{2}}$$

The bound (8.2) follows by multiplying these two factors.

Finally, the estimate (8.3) corresponds to the case when a_C is supported in $|\xi'| > 2^l$ in the above representation. If $T < 2^{-2(k+l)}$ then there are no oscillations in ξ' on the 2^{k+l} scale, and we just use the brute force estimate. For $T > 2^{-2(k+l)}$ the phase is nonstationary in ξ', and we obtain the factor

$$2^{3(k+l)}(1 + 2^{2(k+l)}T)^{-N}$$

$\square$

While the above proposition contains all the oscillatory integral estimates which are needed, it does not apply directly to the frequency localized operators $Op(Ad(O))_{<0}(t, x, D)$ and $Op(Ad(O))_{<0}(D, y, s)$. For that we need to produce similar estimates for the kernels $K_{a,<0}$ of the operators

$$T^a_{<0} = Op(Ad(O))_{<0}(t, x, D)a(D)e^{\pm i(t-s)|D|}Op(Ad(O^{-1}))_{<0}(D, s, y)$$

The transition to such operators is made in the next

Proposition 8.2. *a) Assume that a is a smooth bump on the unit scale. Then the kernel $K^a_{<0}$ satisfies*

$$|K^a_{<0}(t, x; s, y)| \lesssim \langle t - s\rangle^{-\frac{3}{2}}\langle |t - s| - |x - y|\rangle^{-N} \tag{8.4}$$

In addition, the following fixed time bound holds:

$$|K^a_{<0}(t, x; t, y) - \check{a}(x - y)| \leq \epsilon|\log \epsilon| \tag{8.5}$$

b) Let $a = a_C$ be a bump function on a rectangular region C of size $2^k \times (2^{k+l})^3$ with $k \leq l \leq 0$. Then

$$|K^a_{<0}(t, x; s, y)| \lesssim 2^{4k+3l}\langle 2^{2(k+l)}(t - s)\rangle^{-\frac{3}{2}}\langle 2^k(|t - s| - |x - y|)\rangle^{-N} \tag{8.6}$$

c) Let $a = a_C$ be a bump function on a rectangular region C of size $1 \times (2^l)^3$ with $l \leq 0$. Let $\omega \in \mathbb{S}^3$ be at angle l from C. Then we have the characteristic kernel bound

$$\begin{aligned} |K^a_{<0}(t, x; s, y)| &\lesssim 2^{3l}\langle 2^{2l}|t - s|\rangle^{-N}\langle 2^l|x' - y'|\rangle^{-N} \\ t - s &= (x - y)\cdot\omega \end{aligned} \tag{8.7}$$

Proof. a) We represent the action of symbol $Op(O)_{<0}$ by

$$Op(O)_{<0}F(x) = \int m(z)\int e^{i(x-y)\xi}O(x + z, \xi)F(y)O^{-1}(x + z, \xi)dyd\xi\, dz \tag{8.8}$$

where $m(z)$ is an integrable bump function on the unit scale. One proceeds similarly for functions on space-time.

This can be expressed in a concise form using the operators T_z, T_w to represent translation in the space-time directions z, w acting on the variables $t, x.$ and s, y, respectively.

Using this representation for both operators $Op(O)_{<0}$, $Op(O)^*_{<0}$, and denoting $a(z,w)(\xi) = a(\xi)e^{i(\pm|\xi|,\xi)\cdot(z-w)}$, he kernel $K^a_{<0}$ can be expressed in terms of the kernels K^a in the previous proposition, namely

$$K^a_{<0}F(t,x) = \int T_z T_w K^{a(z,w)} F(t,x)\ m(z)m(w)\ dzdw \tag{8.9}$$

To prove the bound (8.4) we use (8.1), together with the additional observation that the implicit constant in (8.1) depends on finitely many seminorms of a (at most 8, to be precise) which we denote by $|||a|||$. Then

$$|||a(z,w)||| \lesssim (1+|z|+|w|)^N$$

However, this growth is compensated by the rapid decay of m, therefore the bound (8.1) for K^a transfers directly to $K^a_{<0}$ in (8.4).

To prove (8.5) we use the same representation as above to write

$$K^a_{<0}F(t,x) - \check{a} * F(t,x) = \int (T_z T_w K^{a(z,w)} - I) F(t,x)\ m(z)m(w)\ dzdw$$

By (7.20) we have

$$|T_z\psi_\pm(t,x,\xi) - T_w\psi_\pm(t,y,\xi))| \lesssim \epsilon \log(1+|z|+|w|+|x-y|)$$

which yields

$$\begin{aligned} |K^a_{<0}(t,x,t,y) - \check{a}(x-y)| &\lesssim \epsilon \int \log(1+|z|+|w|+|x-y|)|m(z)||m(w)|dzdw \\ &\lesssim \epsilon \log(2+|x-y|) \end{aligned}$$

This suffices if $\log(2+|x-y|) \lesssim |\log\epsilon|$. But for larger $|x-y|$ we can use (8.4) directly.

b) Using the representation (8.9), the bound (8.6) follows from (8.2) exactly by the same argument as in case (a).

c) Using the representation (8.9), the same argument also yields the bound (8.1) provided we have the following estimate for K^a:

$$|K^a(t,x,s,y)| \lesssim 2^{3l}\langle 2^{2l}|t-s|\rangle^{-N}\langle 2^l|x'-y'|\rangle^{-N}(1+|(t-s)-(x-y)\cdot\omega|)^{10N}$$

To see that this is true, we consider three cases:

(i) If $|t-s| \lesssim 2^{-2l}$ then (8.2) applies directly.

(ii) If $|t-s| \gg 2^{-2l}$ but $||x-y|-|t-s|| \gtrsim 2^l|x'-y'| + 2^{2l}|t-s|$ then (8.2) still suffices.

(iii) If $|t-s| \gg 2^{-2l}$ and $|(t-s)-(x-y)\cdot\omega|)| \gtrsim 2^l|x'-y'| + 2^{2l}|t-s|$ then (8.2) also applies.

(iv) Finally, if $|t-s| \gg 2^{-2l}$, but $||x-y|-|t-s|| \ll 2^l|x'-y'| + 2^{2l}|t-s|$ and $|(t-s)-(x-y)\cdot\omega|)| \ll 2^l|x'-y'| + 2^{2l}|t-s|$ then we must have $\angle(x-y,\omega) \ll 2^l$, which implies that $\angle(x-y,C) \approx 2^l$. Then (8.3) applies.

$\square$

8.1.1. *Fixed-time L^2 estimates for the gauge transformations.* Here we use the previous theorem to prove three L^2 estimates which correspond to the L^2-part of (6.17), (6.18) as well as that of (6.19). These will also be repeatedly used later in conjunction with the notion of disposability.

Proposition 8.3. *The following fixed time L^2 estimates hold for functions localized at frequency 1, with or without the <0 symbol localization:*

$$Op(Ad(O))_{<0}(t,x,D): \quad L^2 \to L^2, \tag{8.10}$$

$$Op(Ad(O))_{<0}(t,x,D)a(D)Op(Ad(O^{-1}))_{<0}(D,y,s) - a(D): \quad L^2 \to \epsilon^{\frac{N-4}{N}} \log\epsilon \ L^2 \tag{8.11}$$

$$\partial_{x,t} Op(Ad(O))_{<0}(t,x,D): \quad L^2 \to \epsilon L^2 \tag{8.12}$$

Proof. a) By the estimate (8.1) with $s=t$, the TT^* type operator

$$Op(Ad(O))(t,x,D)P_0^2 Op(Ad(O^{-1}))(D,y,t)$$

has an integrable kernel, so it is L^2 bounded. Therefore $Op(Ad(O))(t,x,D)P_0$ and its adjoint are L^2 bounded. To accommodate symbol localizations we observe that

$$Op(Ad(O))_{<k} = \int m_k(z) Op(Ad(T_z O))\, dz$$

where $m(z)$ is an integrable bump function on the 2^{-k} scale and T_z denotes translation in the direction z, with z representing space-time coordinates. Since the wave equation is invariant to translations, the symbol $T_z O$ is of the same type as O and its left and right quantizations are also L^2 bounded. Thus the bound (8.10) follows by integration with respect to z.

b) For the estimate (8.11) we note that the kernel of

$$Op(Ad(O))_{<0}(t,x,D)a(D)Op(Ad(O^{-1}))_{<0}(D,y,s) - a(D)$$

is given by $K^a_{<0}(t,x,t,y) - \check{a}(x-y)$. Combining (8.1) and (8.5) we get

$$|K^a_{<0}(t,x,t,y) - \check{a}(x-y)| \lesssim \min\{\epsilon|\log\epsilon|, |x-y|^{-N}\}$$

The integral of the expression on the right is about $\epsilon^{\frac{N-4}{N}}|\log\epsilon|$, therefore the conclusion follows.

c) By translation invariance we discard the <0 symbol localization, and show that $\partial_{x,t} Op(Ad(O))(t,x,D)P_0$ is L^2 bounded. We have

$$\partial_x Ad(O) = ad(O_{;x})Ad(O)$$

By (7.30) we have $O_{;x} \in \epsilon DL^\infty(L^\infty)$ therefore we can dispose of it and use the L^2 boundedness of $Op(Ad(O))P_0$. □

8.2. **High space-time frequencies in O.** Although $\Psi_{<k}$ is localized at space-time frequencies $<k$, its renormalization counterpart $O_{<k}$ does not share the same property since it is obtained in a nonlinear fashion. Nevertheless, the following result asserts that the high frequency part of $O_{<k}$ does satisfy much better bounds:

Lemma 8.4. *Assume that $1 \leqslant q \leqslant p \leqslant \infty$. Then for $k + C \le l \le 0$ we have :*

$$\| Op(Ad(O_{<k}))_l(t,x;D) \|_{L^p(L^2)\to L^q(L^2)} \lesssim \epsilon 2^{(\frac{1}{p}-\frac{1}{q})k} 2^{5(k-l)}\ , \tag{8.13}$$

This holds for both left and right quantizations.

Proof. For the symbol we iteratively write:

$$\begin{aligned} S_l Ad(O_{<k}) &= 2^{-l} S_l \partial_{x,t}(Ad(O)_{<k}) = 2^{-l} S_l (ad(O_{;(x,t)} Ad(O)_{<k}) \\ &= \ \dots \ = \ 2^{-5l} \prod_{j=1}^{5} (S_l^{(j)} ad(O_{;(x,t)}) \cdot Ad(O)_{<k} \ , \end{aligned}$$

where the product denotes a nested (repeated) application of multiplication by $S_l \partial_t \psi_{<k}$, for a series of frequency cutoffs $S_l^{(j+1)} S_l^{(j)} = S_l^{(j)} \approx S_l$ with expanding widths. Disposing of these translation invariant cutoffs we see that (8.13) follows directly from (7.30).

□

8.3. **Modulation localized estimates.** Our next goal is to show that the fixed time L^2 bounds for $Op(O)$ drastically improve to space-times $L^2(L^2)$ bounds if one selects a fixed "frequency" in the symbol. Precisely, for $k < 0$ we can express the difference

$$Ad(O_{<0}) - Ad(O_{<k}) = \int_k^0 ad(\Psi_h) Ad(O_{<h}) dh$$

where the integrand $Ad(O)_{;h} := ad(\Psi_h) Ad(O_{<h})$, while not exactly localized at frequency 2^h, nevertheless is better behaved both at higher and at lower frequencies. The next result asserts that the output of $Op(Ad(O)_{;h})(t,x,D)$ is better behaved at modulations less than 2^h:

Proposition 8.5. *For $l \leqslant k' \pm O(1)$ one has the fixed frequency estimate:*

$$\| Q_l Op(Ad(O)_{;k'}) Q_{<0} P_0 \|_{N^* \to X_1^{0,\frac{1}{2}}} \ \lesssim \ 2^{\delta(l-k')} \epsilon \ . \tag{8.14}$$

In particular summing over all (l, k') with $l \leqslant k$ and $k - O(1) \leqslant k'$ for a fixed $k \leqslant 0$ yields:

$$\| Q_{<k}(Op(Ad(O_{<0})) - Op(Ad(O_{<k-C})) Q_{<0} P_0 \|_{N^* \to X_1^{0,\frac{1}{2}}} \ \lesssim \ \epsilon \ . \tag{8.15}$$

Proof of Proposition 8.5. We proceed in a series of steps, where we consider successive modulation scenarios.

Step 1:(*High modulation input*) First we estimate the contribution of the dyadic piece $Q_k Op(Ad(O)_{;k'}) Q_{\geqslant k-C} P_0$ to line (8.14). Using the $X_\infty^{0,\frac{1}{2}}$ bounds for the input, it suffices to prove the estimate:

$$\| Q_k Op(Ad(O)_{;k'}) P_0 \|_{L^2(L^2) \to L^2(L^2)} \ \lesssim \ 2^{\frac{1}{5}(k-k')} \epsilon \ .$$

By Sobolev estimates in $|\tau| \pm |\xi|$, this reduces to the bound:

$$\| Op(Ad(O)_{;k'}) P_0 \|_{L^2(L^2) \to L^{\frac{10}{7}}(L^2)} \ \lesssim \ 2^{-\frac{1}{5}k'} \epsilon \ .$$

Recalling that $Op(Ad(O)_{;k'})$ has symbol $ad(\Psi_{k'}) Ad(O_{<k'})$, it suffices to use the L^2 boundedness for $Op(O_{<k'})$ and the $L^5 L^\infty$ disposability bound for $\Psi_{k'}$.

Step 2:(*Main decomposition for low modulation input*) Now we estimate the expression $Q_k Op(Ad(O)_{;k'})Q_{<k-C}P_0 u$. First expand the untruncated group elements as follows:

$$
\begin{aligned}
Ad(O)_{;k'} &= ad(\Psi_{k'})Ad(O_{<k-C}) + \int_{k-C}^{k'} ad(\Psi_{k'})ad(\Psi_l)Ad(O_{<k-C})dl \\
&\quad + \int_{k-C'}^{k'}\int_{l'}^{k'} ad(\Psi_{k'})ad(\Psi_l)ad(\Psi_{l'})Ad(O_{<l'})dldl' \\
&= \mathcal{L} + \mathcal{Q} + \mathcal{C}.
\end{aligned} \tag{8.16}
$$

We will estimate the effect of each of these terms separately.

Step 3:(*Estimating the linear term* $\mathcal{L}$) The factor $ad(\Psi_{k'})$ in $\mathcal{L}$ is well localized both in frequency and modulation. While not exactly localized, the second factor $Ad(O_{<k-C})$ is to the leading order localized at frequency and modulation $\leq k - C/2$, with more regular and decaying tails at larger frequencies and modulations. The geometry of the bilinear wave interactions, on the other hand, requires us to estimate differently the contribution of $ad(\Psi_{k'})$ depending on its modulation relative to 2^k. To account for both considerations above, we split the term $\mathcal{L}$ as follows:

$$
\mathcal{L} \ = ad(\Psi_{k'})S_{<k-4}Ad(O_{<k-C}) + ad(\psi_{k'})S_{>k-4}Ad(O_{<k-C}) \tag{8.17}
$$

Step 3a:(*Estimating the principal linear term in* $\mathcal{L}$) For the first term on RHS of line (8.17) it suffices to show the general estimate:

$$
\| Q_k Op(ad(\Psi_{k'})b_{<k-4})Q_{<k-C}P_0 \|_{L^\infty(L^2)\to L^2(L^2)} \lesssim \epsilon\, 2^{-\frac{1}{2}k+\frac{1}{4}(k-k')} \sup_t \| B_{<k-4}(t) \|_{L^2\to L^2} \tag{8.18}
$$

for $k' \geqslant k$, and for symbols $b(x,\xi)_{<k-4}$ with sharp frequency and modulation localization and with either the left or right quantization. The geometry of the bilinear wave interactions requires us to estimate differently the contribution of $ad(\Psi_{k'})$ depending on its modulation relative to 2^k. Thus we will consider three cases:

Step 3a(i):(*The contribution of* $Q_{<k}\Psi_{k'}$) In this case the modulation of the output determines the angle θ between the spatial frequencies of $\Psi_{k'}(x,\xi)$ and the spatial frequency of the input, which is $\theta \sim 2^{\frac{1}{2}(k-k')}$. Since this is also the angle with ξ, we may restrict the symbol of $\Psi_{k'}$ to $\psi_{k'}^{(\theta)}$ for which the estimate (8.18) follows immediately from (7.5) and summing over (7.13).

Step 3a(ii):(*The contribution of* $Q_k\Psi_{k'}$) In this case one of the inputs has the same modulation as the output, so we only get a bound from above on the angle θ, namely $\theta \lesssim 2^{\frac{1}{2}(k-k')}$. However, instead of (7.13), which looses at small angles, we can take advantage of the fixed modulation to use (7.14), which gains at small angles.

Step 3a(iii):(*The contribution of* $Q_{>k}\Psi_{k'}$) In this case one of the inputs has high modulation, say $2^{k'+2j'}$ with $(k-k')/2 < j' \leq 0$. This determines the angle θ to be $\theta \approx k' + j'$. Then we can use again (7.14).

Step 3b:(*Estimating the frequency truncation error in* $\mathcal{L}$) For the second term on RHS of line (8.17) we use (7.16) for $\psi_{k'}$ with $p = 6$ combined with (8.13) with $(p_2, q) = (6, 3)$.

Step 4:(*Estimating the quadratic term $\mathcal{Q}$*) We follow a similar procedure to **Step 3** above. First split $S_{<k-4}Ad(O_{<k-C}) + S_{>k-4}Ad(O_{<k-C})$. For the second term one can proceed as in **Step 4b** above using (7.16), (8.13), and (7.3).

Therefore we only need to consider the effect of the first term, for which we will prove the trilinear bound:

$$\begin{aligned} (8.19)\quad & \| Q_k \cdot Op(ad(\Psi_{k'}ad(\Psi_l)b_{<k-4})(t,x;D) \cdot Q_{<k-C}P_0 \|_{L^\infty(L^2)\to L^2(L^2)} \\ & \qquad \lesssim \epsilon^2\, 2^{-\frac{1}{2}k}2^{\frac{1}{4}(k-k')}2^{\frac{1}{6}(k-l)} \sup_t \| B_{<k-4}(t) \|_{L^2\to L^2} \ , \end{aligned}$$

for $k' \geqslant l \geqslant k$. We decompose the symbol $ad(\Psi_{k'})ad(\Psi_l)$ in terms of the angles,

$$\sum_{\theta \gtrsim 2^{\frac{1}{2}(k-k')}} ad(\Psi^{(\theta)}_{k'})ad(\Psi_l) + \sum_{\substack{\theta \gtrsim 2^{\frac{1}{2}(k-k')} \\ \theta' \ll 2^{\frac{1}{2}(k-l)}}} ad(\Psi^{(\theta)}_{k'})ad(\Psi^{(\theta')}_l) + \sum_{\substack{\theta \ll 2^{\frac{1}{2}(k-k')} \\ \theta \ll 2^{\frac{1}{2}(k-l)}}} ad(\Psi^{(\theta)}_{k'})ad(\Psi^{(\theta')}_l) \ = \ T_1 + T_2 + T_3 \ .$$

For the term T_1 put the first factor in $DL^3(L^\infty)$ and the second in $DL^6(L^\infty)$. This gives us dyadic terms in LHS(8.19)(T_1) $\sim 2^{-\frac{1}{2}k}2^{\frac{1}{4}(k-l)}2^{\frac{1}{6}(k-l')}$. For the term T_2 do the opposite, which yields a similar bound. Finally, for the term T_3 a frequency modulation analysis shows that at least one of the two factors has modulation $\geq k$. Then we use (7.14) to place that factor in $DL^2(L^\infty)$ and simply bound the remaining factor in $DL^\infty L^\infty$.

Step 5:(*Estimating the cubic term $\mathcal{C}$*) In this case we can gain $2^{\frac{1}{6}(k-k')}$ directly through the use of (7.16) and three $DL^6(L^\infty)$. Further details are left to the reader.

□

8.4. **The N_0 and N_0^* bounds in** (6.17)**,** (6.18) **and** (6.19)**.** We are now ready to conclude the proof of the first part of Theorem 3.

Proof of (6.17) *for $Z = N_0, N_0^*$.* By duality it suffices to prove the N_0^* bound for both the left and the right calculus. The $L^\infty L^2$ bound follows from the fixed time L^2 bound. The $X^{0,1}_\infty$ bound is also straightforward when we go from high to low modulation. It remains to consider the case of low modulation input and high modulation output. Precisely, we need to show that

$$\|Q_k Op(Ad(O))Q_{<k-C}P_0\|_{L^\infty L^2 \to L^2} \lesssim \epsilon 2^{-\frac{k}{2}} \tag{8.20}$$

From here on, we specialize to the left calculus. By (8.15), it remains to estimate

$$\|Q_k Op(Ad(O_{<k-C}))Q_{<k-C}P_0\|_{L^\infty L^2 \to L^2 L^2} \lesssim \epsilon 2^{-\frac{k}{2}}$$

Here we can harmlessly replace $Ad(O_{<k-C})$ by $S_{>k-4}Ad(O_{<k-C})$. But then we can conclude using (8.13).

To prove (8.20) for the right calculus, we use duality to switch to the left calculus bound

$$\|Q_{<k-C} Op(Ad(O))Q_{<k}P_0\|_{L^2 \to L^1 L^2} \lesssim \epsilon 2^{-\frac{k}{2}} \tag{8.21}$$

Then we can conclude the proof in the same manner as before. □

Proof of (6.18) *for $Z = N_0, N_0^*$.* Here we repeat the above analysis with $Ad(O)$ replaced by $\partial_t(Ad(O)) = ad(O_{;t})Ad(O)$. We remark that

$$\partial_h \partial_t(Ad(O))) = ad(\partial_t(\Psi_h))Ad(O) + ad(\Psi_h)ad(O_{;t})Ad(O)$$

and all terms above are of the same form as above, possibly with $Ad(O))$ harmlessly replaced by $ad(O_{;t})Ad(O)$. $\square$

Proof of (6.19) *for* $Z = N_0, N_0^*$. By duality it suffices to consider the case $Z = N_0^*$. In view of the L^2 bound proved earlier, it suffices to show that

$$\|Q_k Op(Ad((O))Op(Ad(O))^* Q_{<k-C}\|_{L^\infty L^2 \to L^2} \lesssim \epsilon 2^{-\frac{k}{2}}$$

But this is a consequence of two bounds,

$$\|Q_k Op(Ad(O)) Q_{<k-C}\|_{L^\infty L^2 \to L^2} \lesssim \epsilon 2^{-\frac{k}{2}}$$

and

$$\|Q_{>k-C/2} Op(Ad(O))^* Q_{<k-C}\|_{L^\infty L^2 \to L^2} \lesssim \epsilon 2^{-\frac{k}{2}}$$

both of which follow from (8.20). $\square$

8.5. **Strichartz and null frame norm estimates.** Here we briefly outline how to prove the bound (6.21). In fact, the argument for this bound follows exactly like the proof of (83) in section 11 of [16]. One replaces (114) in [16] by the L^2-boundedness of the operators $Op(Ad(O_\pm)_{<k})(t,x,D)$, the dispersive bounds (108), (110) in [16] by the bounds (8.4), (8.6), and the bound (118) in [16] by (8.15).

9. Error estimates

Here we again simplify notation by writing $O_{<0} = O$. The goal of this section is to consider the conjugation error

$$E = \square^p_{A_{<0}} Op(Ad(O)) - Op(Ad(O))\square$$

and prove the bound (6.20) in Theorem 3.

Commuting $\square$ we have

$$\begin{aligned} E =& 2Op(ad(A_{<0,\alpha})Ad(O))\partial^\alpha + 2Op(ad(A_{<0,\alpha})ad(O^{;\alpha})Ad(O)) + 2Op(ad(O_{;\alpha})Ad(O))\partial^\alpha \\ &+ Op(ad(\partial^\alpha O_{;\alpha})Ad(O)) + Op(ad(O_{;\alpha})ad(O^{;\alpha})Ad(O)) \\ =& \ 2Op(ad(A_{<0,\alpha} + \Psi_{<0,\alpha})Ad(O))\partial^\alpha \\ &+ 2Op(ad(O_{;\alpha} - \Psi_\alpha)Ad(O))\partial^\alpha \\ &+ 2Op(ad(A_{<0,\alpha})ad(O^{;\alpha})Ad(O)) + Op(ad(O_{;\alpha})ad(O^{;\alpha})Ad(O)) \\ &+ Op(ad(\partial^\alpha O_{;\alpha})Ad(O)) \\ =& \ E_1 + E_2 + E_3 + E_4 \end{aligned}$$

Here the main difficulty is to estimate the term E_1, which not only contains the input of $A^{pert,\pm}_{j,<0}$ but also the full input from A_0. We carry out a good portion of the analysis in Section 9.1, modulo a single interaction scenario which is more extensive and requires more than the S norm of A ; this is relegated to the last section 10. The remaining terms E_2, E_3 and E_4 are dealt with in Section 9.2. These are more in line with previous estimates, and only require the S^1 norm of A_x.

9.1. **The estimate for E_1.** We recall that

$$\Psi_{k,+}(t,x,\xi) \;=\; L_-^\omega \Delta_{\omega^\perp}^{-1}(A_{j,k}^{main} \cdot \omega_j), \qquad A_{j,k}^{main} = \Pi_{>\delta k}^\omega \Pi_{cone}^\omega A_{j,k}$$

where

$$L_-^\omega = \partial_t - \omega \nabla_x$$

Replacing the operator D_t by $-|D_x|$ we produce a first error, namely

$$Op(ad(A_0 + \partial_0 \Psi)Ad(O))(D_t + |D_x|)$$

which is easily dealt with using DL^2L^∞ disposability bounds for A_0 and $\partial_t \Psi$. We are left with

$$E_1 = Op(ad(A_j \cdot \omega + A_0 + L_+^\omega \Psi^+)Ad(O))$$

Now we use

$$L_+^\omega L_-^\omega \Delta_{\omega^\perp}^{-1} = \Box \Delta_{\omega^\perp}^{-1} - 1$$

to write

$$G := A_j \cdot \omega + A_0 + L_+^\omega \Psi^+ = G_{cone} + G_{null} + G_{out}$$

where

$$\begin{aligned} G_{cone} &= \Box \Delta_{\omega^\perp}^{-1} \Pi_{>\delta k}^\omega A_{j,cone}\omega_j + \Pi_{<\delta k}^\omega A_{j,cone}\omega_j + A_{0,cone} \\ G_{null} &= A_{j,null}\omega_j + A_{0,null} \\ G_{out} &= A_{j,out}\omega_j + A_{0,out} \end{aligned}$$

We seek to prove that $Op(ad(G)Ad(O)) : N^* \to N$. We do this in two stages. First we will show that we can dispense with O, and simply prove that

$$Op(ad(G)) : S_0 \to N \tag{9.1}$$

Since $Op(Ad(O))$ is bounded from $S_0^\sharp$ into S^0, in order to achieve this it suffices to show that

$$Op(ad(G)Ad(O)) - Op(ad(G))Op(Ad(O)) : N^* \to N \tag{9.2}$$

This latter bound will not follow immediately from pdo calculus, since G is not smooth with respect to ξ on the unit scale. Instead, our strategy will be to first peel off a contribution which is bad from the perspective of pdo calculus but has a good decomposable structure. For this we consider the pieces $G_h^{(\theta)}$ of G, which are localized at frequency 2^h and angle θ with respect to ω. In view of the bounds (7.7) and (7.13) they satisfy

$$\|G_h^{(\theta)}\|_{DL^2L^\infty} \lesssim \theta^{\frac{3}{2}} 2^{\frac{h}{2}}$$

These symbols are smooth in ξ on the θ scale, so it is natural to match them against symbols which are smooth in x on the θ^{-1} scale. Thus, let h_θ be defined by $2^{h_\theta} = \theta$. Then we decompose the above difference as

$$\begin{aligned} D_h^{(\theta)} &= Op(ad(G_h^{(\theta)})Ad(O)) - Op(ad(G_h^{(\theta)}))Op(Ad(O)) \\ &= \int_{h_\theta}^0 Op(ad(G_h^{(\theta)})ad(\Psi_k)Ad(O_{<k}))dk \\ &\quad - \int_{h_\theta}^0 Op(ad(G_h^{(\theta)})Op(ad(\Psi_k)Ad(O_{<k}))dk \\ &\quad + Op(ad(G_h^{(\theta)})Ad(O_{<h_\theta})) - Op(ad(G_h^{(\theta)}))Op(Ad(O_{<h_\theta})) \end{aligned}$$

For the first term, decomposable estimates show

$$\|Op(ad(G_h^{(\theta)})ad(\Psi_k)Ad(O_{<k}))\|_{L^\infty L^2 \to L^1 L^2} \lesssim \|G_h^{(\theta)}\|_{DL^2L^\infty} \|\Psi_k\|_{DL^2L^\infty} \lesssim \theta^{\frac{3}{2}} 2^{\frac{h}{2}} 2^{(-\frac{1}{2}-\delta)k}$$

which is favorable in view of the range $\theta < 2^k < 1$. A similar argument applies for the second term. For the third term, instead, we can use the pdo calculus. For $|\alpha| \geq 1$ we have

$$\|\partial_\xi^\alpha G_h^{(\theta)}\|_{DL^2L^\infty} \leq c_\alpha \theta^{-|\alpha|} \theta^{\frac{3}{2}} 2^{\frac{h}{2}}$$

while (using Lemma 7.9)

$$\|\partial_x^\alpha Ad(O_{<h_\theta})\|_{L^\infty L^2 \to L^2 L^2} \lesssim \theta^{|\alpha| - \frac{1}{2} - \delta}$$

It follows that

$$\|Op(ad(G_h^{(\theta)})Ad(O_{<h_\theta})) - Op(ad(G_h^{(\theta)}))Op(Ad(O_{<h_\theta}))\|_{L^\infty L^2 \to L^1 L^2} \lesssim \theta^{\frac{1}{2}} 2^{\frac{h}{2}} \theta^{\frac{1}{2}-\delta}$$

which again suffices. Thus the bound (9.2) is proved. We now return to (9.1).

Corresponding to the partition of G into three parts we will also partition

$$E_1 = E_{1,cone} + E_{1,null} + E_{1,out}$$

In this section we will estimate $E_{1,cone}$ and $E_{1,out}$. We will postpone the bound for $E_{1,null}$ for the next section.

The bound for $E_{1,cone}$. The redeeming feature of $E_{1,cone}$ is that the modulation localization and the angle are mismatched and that forces a large modulation on either the input or the output. Precisely, consider the G_{cone} component $G_{cone,k}^{(\theta)}$ at frequency k and angle θ. Then $G_{cone,k}^{(\theta)}$ has modulation at most $2^k\theta^2$, whereas either the input or the output must have modulation at least $2^k\theta^2$. Hence we can use the L^2 norm for either the input or the output, therefore it suffices to have L^2L^∞ disposability for the terms in G_{cone}. Precisely, we obtain

$$\|E_{1,cone}\|_{N^* \to N} \lesssim \sum_{k<0} \sum_{\theta<1} \theta^{-1} 2^{-k/2} \|G_{cone,k}^{(\theta)}\|_{DL^2L^\infty}$$

The nontrivial business is to insure summation. In the second term in G_{cone} we gain from the angle, and thus also in k. In the first term we use disposability derived from the L^2 bound for $\Box A_k$ therefore we gain in angle, and ℓ^1 summation in k. Same for the third term.

The bound for $E_{1,out}$.. Again the modulation localization and the angle are mismatched and that forces a large modulation on either the input or the output. Precisely, consider the G_{cone} component $Q_{k+2j}G_{out,k}^{(\theta)}$ at frequency k and angle θ. Then $G_{cone,k}^{(\theta)}$ has modulation $2^{k+2j} \geq 2^k\theta^2$, whereas either the input or the output must have modulation at least comparable. Hence we can again use the L^2 norm for either the input or the output, therefore it suffices to have L^2L^∞ disposability for the terms in G_{cone}. We obtain

$$\begin{aligned}\|E_{1,cone}\|_{N^* \to N} &\lesssim \sum_{k<0} \sum_{j<0} \sum_{\theta<2^j} 2^{-(k+2j)/2} \|Q_{k+2j} G_{out,k}^{(\theta)}\|_{DL^2L^\infty} \\ &\lesssim \sum_{k<0} \sum_{j<0} \sum_{\theta<2^j} 2^{-\frac{(k+2j)}{2}} \theta \theta^{\frac{3}{2}} 2^{2k} \big\|P_k Q_{k+2j} A_x\big\|_{L^2L^2} + 2^{-(k+2j)/2} \theta^{\frac{3}{2}} 2^{2k} \big\|P_k A_0\big\|_{L^2L^2}\end{aligned}$$

The first term comes from A_j and the second from A_0. The latter has ℓ^1 dyadic summation, while for the former we use Proposition 5.4.

The bound for $E_{1,null}$. We can dispense with the case when either the input or the output have high modulation ($\gtrsim 2^k\theta^2$, where k, θ stand for the frequency, respectively the angle of A) as in the case of $E_{1,cone}$. We are then left with the expression

$$\mathcal{H}^* Op(ad(A_{\alpha,<0}))\partial^\alpha C$$

The bound for this expression is stated in the following lemma, whose proof is relegated to the next section:

Lemma 9.1. *Suppose that A has S^1 norm at most ϵ and solves the YM-CG equation in a time interval I. Extend A_x to a free wave outside I, and A_0 by 0. Then for C at frequency 1 we have the estimate*

$$\|\mathcal{H}^* Op(ad(A_{\alpha,<0}))\partial^\alpha C\|_N \lesssim \epsilon \|C\|_S \tag{9.3}$$

9.2. **The estimates for E_2, E_3 and E_4.** For these terms we can directly use the decomposability bounds bounds on Ψ and O in the previous sections. We consider them successively.

9.2.1. *The E_2 term.* For the second term in the error we recall that

$$\partial_h O_{;\alpha} = \Psi_{h,\alpha} + [\Psi_h, O_{;\alpha}]$$

Thus, repeatedly expanding the symbol $ad(O_{;\alpha} - \Psi_\alpha)Ad(O)$ (by means of (6.13)) with respect to h, we are left with an integral with respect to decreasing h's of expressions of the form

$$ad(\Psi_{h_1})ad(\partial_\alpha \Psi_{h_2})Ad(O_{<h_2}), \quad \cdots \quad ad(\Psi_{h_1})\cdots ad(\Psi_{h_5})ad(\partial_\alpha \Psi_{h_6})Ad(O_{<h_6})$$

plus a final remainder term

$$ad(\Psi_{h_1})\cdots ad(\Psi_{h_5})ad(O_{<h_6;\alpha})Ad(O_{<h_6})$$

with possibly changed order of factors.

For the sixth-linear terms we use DL^6L^∞ bounds for all factors (in particular we need this for $O_{<h_6;\alpha}$ with no loss; $DL^\infty L^\infty$ would also do by reiterating once more).

For the lower order expressions we are in the same situation as in the MKG case, with the critical difference that the Ψ's may now have nonzero modulations. We discuss the second order term, as all higher order terms are similar.

$$ad(\Psi_{h_1})ad(\partial_\alpha \Psi_{h_2})Ad(O_{<h_2})\partial^\alpha$$

Replacing ∂^0 by $-|\xi|$ (with a better error) this becomes

$$ad(\Psi_{h_1})ad(L^+\Psi_{h_2})Ad(O_{<h_2})|\xi|$$

and doing the symbol computation, this has the form

$$D_2 = ad(\Psi_{h_1})ad(A^{main}_{j,h_2}\omega_j)Ad(O_{<h_2})|\xi|$$

Now we do an angle/modulation analysis. We begin with angles, and denote by θ_1, θ_2 the two angles. Then by (7.13) and (7.7) we can first estimate

$$\|D_2\|_{L^\infty L^2 \to L^1 L^2} \lesssim \|\Psi^{(\theta^1)}_{h_1}\|_{DL^2L^\infty} \|A^{main,(\theta^2)}_{j,h_2}\cdot \omega\|_{DL^2L^\infty} \lesssim 2^{(h_2-h_1)/2}\theta_1^{-\frac{1}{2}}\theta_2^{\frac{3}{2}}$$

This is favorable if $2^{h_2}\theta_2^2 \gtrsim 2^{h_1}\theta_1^2$. If this is not the case, then either one of the factors or the input or the output must have modulation at least as large as $2^{h_1}\theta_1^2$. This cannot be the case for $A^{main,(\theta^1)}_{j,h_2}$ by definition, so we have three scenarios to consider:

a) High modulation input. Then by (7.13) and (7.7) we have

$$\|D_2 B\|_{L^1L^2} \lesssim 2^{-h_2/2}\theta_2^{-1}\|\Psi_{h_1}^{(\theta^1)}\|_{DL^6L^\infty}\|A_{j,h_2}^{main,(\theta^2)}\cdot\omega\|_{DL^3L^\infty}\|B\|_{N^*} \lesssim 2^{(h_2-h_1)/6}\theta_2^{2-\frac{1}{6}}$$

which suffices.

b) High modulation output where we have exactly the same bound.

c) High modulation on Ψ_1. Then we can use (7.14) for its DL^2L^∞ bound.

9.2.2. *The term E_3.* In this term we have high frequencies to spare. For $[O_{;\alpha},[O_;^\alpha, Op(O)\cdot]]$ we need some mild L^2L^∞ disposability estimate for $O_{;\alpha}$. Similarly, for A_α we can use an L^2L^∞ bound.

9.2.3. *The term E_4.* For $ad(\partial^\alpha O_{;\alpha})Ad(O)$ we expand in h

$$ad(\partial^\alpha O_{;\alpha})Ad(O) = \int_{-\infty}^0 (ad(\Box\Psi_h) + \partial^\alpha[\Psi_h, O_{<h;\alpha}] + ad(\partial^\alpha O_{<h;\alpha})ad(\Psi_h))Ad(O_{<h})dh$$

For the second and third term we use L^2L^∞ disposability for Ψ_h and $O_{<h;\alpha}$, with room to spare. For the first term, Consider the component $ad(\Box\Psi_h^{(\theta)})$ at angle θ and reexpand with $2^{h_\theta} = \theta^2 2^h$:

$$ad(\Box\Psi_h^{(\theta)})Ad(O_{<h}) = ad(\Box\Psi_h^{(\theta)})Ad(O_{<h_\theta - C}) + \int_{h_\theta - C}^h ad(\Box\Psi_h^{(\theta)})ad(\Psi_{h_1})Ad(O_{<h_1})dh_1$$

For the integrand we can use two DL^2L^∞ bounds to estimate

$$\|\Box\Psi_h^{(\theta)}\|_{DL^2L^\infty}\|\Psi_{h_1}\|_{DL^2L^\infty} \lesssim \theta^{\frac{3}{2}}2^{\frac{3h}{2}}2^{-(\frac{1}{2}+\delta)h_1}$$

which is favorable due to the range of h_1. For the leading term, using (8.13), we replace $Ad(O_{<h_\theta-C})$ by $S_{<h_\theta-4}Ad(O_{<h_\theta-C})$. At this stage we are left with the operator

$$Op(ad(\Box\Psi_h^{(\theta)})S_{<h_\theta-4}Ad(O_{<h_\theta-C}))$$

Given the frequency localization of $\Psi_h^{(\theta)}$, the space-time frequency interaction analysis shows that either the input or the output must have modulations at least $2^h\theta^2$. Then we can conclude using the DL^2L^∞ disposability of $\Box\Psi_h^{(\theta)}$ in (7.15).

□

10. Trilinear forms and the second null structure

Here we prove Lemma 9.1 and Lemma 5.6, which we restate for convenience:

Lemma 10.1. *a) Suppose that A has S^1 norm at most ϵ and solves the YM-CG equation in a time interval I. Extend A_x to a free wave outside I, and A_0 by 0. Then for C_k at frequency 2^k we have the estimate*

$$\|\mathcal{H}^*[A_{\alpha,<k}, \partial^\alpha C_k]\|_N \lesssim \epsilon\|C_k\|_S \tag{10.1}$$

b) Suppose in addition that $B \in S^s$ solves the linearized equation (1.10) *in a time interval I. Extend B_j outside I as free waves, and B_0 by zero. Then for $s < 1$, close to 1 we have the global estimate*

$$\|\mathcal{H}^*[B_{\alpha,<k}, \partial^\alpha C_k]\|_{N^{s-1}} \lesssim \epsilon\|B\|_{S^s}\|C_k\|_{S^1} \tag{10.2}$$

The proofs for the two parts are quite similar, and hinge on a double null structure in the main trilinear expression arising when one replaces the first factor in the expressions above with the solutions of the corresponding $\Box$ equation for A_x and B_x, respectively the Δ equation for A_0 and B_0.

Proof of Lemma 10.1. a) To better frame the question, denote by 2^h, θ the frequency, respectively the angle of A. Then the $\mathcal{H}^*$ operator selects the cases where both the input and the output are at modulation less than $2^h\theta^2$.

Our first tool here is to use the Z norm bounds (5.5), (5.7). To bound (most of) A_x and A_0 we use their equations (1.6), respectively (1.7). We claim that the following hold:

$$\begin{aligned} \|\Box A_j - \mathcal{H}\mathbf{P}[A_i, \chi_I \partial_j A_i]\|_{\ell^1 \Box Z} &\lesssim \epsilon^2 \\ \|\Delta A_0 - \mathcal{H}[A_i, \chi_I \partial_0 A_i]\|_{\ell^1 \Box^{\frac{1}{2}} \Delta^{\frac{1}{2}} Z} &\lesssim \epsilon^2 \end{aligned} \tag{10.3}$$

For this we consider all other terms in the equations for A_j and A_0, which we recall here:

$$\Box A_j = \mathbf{P}\left([A^\alpha, \partial_j A_\alpha] - 2[A^\alpha, \partial_\alpha A_j] - [\partial_0 A_0, A_j] - [A^\alpha, [A_\alpha, A_j]]\right)$$

$$\Delta A_0 = [A_j, \partial_0 A_j] - 2[A_j, \partial_j A_0] - [A_j, [A_j, A_0]]$$

Here we seemingly pay a price for working in an interval I, as both right hand sides need to be multiplied by the characteristic function χ_I of I. However, this turns out to be harmless, because we can always place χ_I on the differentiated factor, and still retain the use of the S norm.

(i) Cubic terms A^3. These are placed in $\ell^2 L^1 L^2$ which suffices by (5.9). (we do need to gain ℓ^1 summability in k).

(ii) $[A_j, \partial_j A_0]$ and $[\partial_0 A_0, A_j]$. are also in $\ell^1 L^1 L^2$ by using $L^2 \dot{H}^{\frac{1}{2}}$ for ∇A_0 and $L^2 L^6$ for A_j.

(iii) The term $[A_0, \partial_0 A_j]$. The low-high case is the worst, but even then we can use Strichartz to produce $L^1 L^\infty$.

(iv) High-low interactions in the quadratic terms $A_j \nabla A_k$. This is where we use (5.10).

(v) High-high interactions in $[A_j, \partial_j A_k]$. Here we can take the derivative out and estimate as in the high-low case via (5.10)..

(vi) High-high interactions in $[A_j, \partial_\alpha A_j]$. with at least one high modulation Here by estimating one factor in L^2 we can gain in terms of high frequencies, see (5.12).

This concludes the proof of (10.3). In view of (5.5), (5.7), this leaves us with one remaining case:

(Final case) High-high interactions in $[A_j, \partial_\alpha A_j]$ with two low modulations. Here we need to combine the $\Box^{-1} A_j$ and $\Delta^{-1} A_0$ contributions in order to gain an additional cancellation.

Omitting the frequency and modulation localizations, the expression is as follows:

$$\begin{aligned} L =& (\Box^{-1}\mathbf{P}[A_j, \partial_k A_j]\partial_k F + \Delta^{-1}[A_j, \partial_0 A_j])|D|F \\ =& \Box^{-1}[A_j, \partial_k A_j]\partial_k F - \frac{\partial_k \partial_i}{\Box\Delta}[A_j, \partial_i A_j]\partial_k F - \Delta^{-1}[A_j, \partial_0 A_j]\partial_0 F \\ =& \Box^{-1}[A_j, \partial_\alpha A_j]\partial^\alpha F - \frac{\partial_k \partial_i}{\Box\Delta}[A_j, \partial_i A_j]\partial_k F + \frac{\partial_0^2}{\Box\Delta}[A_j, \partial_0 A_j]\partial_0 F \\ =& \Box^{-1}[A_j, \partial_\alpha A_j]\partial^\alpha F - \frac{\partial_\alpha \partial_i}{\Box\Delta}[A_j, \partial_i A_j]\partial^\alpha F - \frac{\partial_0 \partial_i}{\Box\Delta}[A_j, \partial_i A_j]\partial_0 F + \frac{\partial_0^2}{\Box\Delta}[A_j, \partial_0 A_j]\partial_0 F \\ =& \Box^{-1}[A_j, \partial_\alpha A_j]\partial^\alpha F - \frac{\partial_\alpha \partial_i}{\Box\Delta}[A_j, \partial_i A_j]\partial^\alpha F - \frac{\partial_0 \partial_\alpha}{\Box\Delta}[A_j, \partial^\alpha A_j]\partial_0 F \end{aligned}$$

The estimate for this term is exactly the trilinear bound in [16], see (136) - (138) in Theorem 12.1 there.

b) This is similar to the proof in part (a), with two differences:

i) There is an additional gain in the low frequency input, which eliminates any need to control ℓ^1 norms.

ii) There is a small additional loss in high-high interactions in $\Box A_x$ and ΔA_0. However, this is harmless as in all cases we have a small high frequency gain (including, notably, the trilinear case).

□

References

[1] Ioan Bejenaru, Sebastian Herr, *The cubic Dirac equation: Small initial data in $H^{\frac{1}{2}}(\mathbb{R}^2)$*, ArXiv e-prints, arXiv:1501.06874

[2] Ioan Bejenaru, Sebastian Herr, *On global well-posedness and scattering for the massive Dirac-Klein-Gordon system*, ArXiv e-prints, arXiv:1409.1778

[3] Hajer Bahouri and Patrick Gérard, *High frequency approximation of solutions to critical nonlinear wave equations*, Amer. J. Math. **121** (1999), no. 1, 131–175. MR 1705001 (2000i:35123)

[4] Yvonne Choquet-Bruhat and Demetrios Christodoulou, *Existence of global solutions of the Yang-Mills, Higgs and spinor field equations in $3+1$ dimensions*, Ann. Sci. de l'E.N.S., 4eme serie, tome 14, no.4(1981), p. 481 – 506.

[5] Scipio Cuccagna, *On the local existence for the Maxwell-Klein-Gordon system in R^{3+1}* Comm. PDE **24** (1999), no. 5-6, 851–867

[6] Douglas M. Eardley and Vincent Moncrief, *The global existence of Yang-Mills-Higgs fields in 4-dimensional Minkowski space. I. Local existence and smoothness properties*, Comm. Math. Phys. **83** (1982), no. 2, 171–191. MR 649158 (83e:35106a)

[7] ______, *The global existence of Yang-Mills-Higgs fields in 4-dimensional Minkowski space. II. Completion of proof*, Comm. Math. Phys. **83** (1982), no. 2, 193–212. MR 649159 (83e:35106b)

[8] ______, *Global well-posedness, scattering and blow-up for the energy-critical focusing non-linear wave equation*, Acta Math. **201** (2008), no. 2, 147–212. MR 2461508 (2011a:35344)

[9] Sergiu Klainerman and Matei Machedon, *On the Maxwell-Klein-Gordon equation with finite energy*, Duke Mathematical Journal (1994).

[10] Sergiu Klainerman and Matei Machedon, *Finite energy solutions for the Yang-Mills equations in $\mathbb{R}^{3+1}$*, Annals of Mathematics, Vol. 142(1995), 39 – 119.

[11] Sergiu Klainerman and Daniel Tataru, *On the optimal local regularity for Yang-Mills equations in $\mathbf{R^{4+1}}$*, Journal of the American Mathematical Society (1999).

[12] Herbert Koch, Daniel Tataru, and Monica Vişan, *Dispersive equations and nonlinear waves*, Springer, 2014.

[13] Joachim Krieger, Jonas Luhrmann, *Concentration Compactness for the Critical Maxwell-Klein-Gordon Equation* ArXiv e-prints, arXiv:1503.09101
[14] Joachim Krieger and Wilhelm Schlag, *Concentration compactness for critical wave maps*, EMS Publishing House, 2009.
[15] Joachim Krieger and Jacob Sterbenz, *Global Regularity for the Yang-Mills Equations on High Dimensional Minkowski Space*, arXiv.org (2005).
[16] Joachim Krieger, Jacob Sterbenz, and Daniel Tataru, *Global well-posedness for the Maxwell-Klein Gordon equation in 4+1 dimensions. Small energy*, arXiv.org (2012), Duke Mathematical Journal, to appear.
[17] Matei Machedon and Jacob Sterbenz, *Almost optimal local well-posedness for the* $(3+1)$*-dimensional Maxwell-Klein-Gordon equations*, Journal of the American Mathematical Society (2004).
[18] Vincent Moncrief, *Global existence of Maxwell-Klein-Gordon fields in* $(2+1)$*-dimensional spacetime*, J. Math. Phys. **21** (1980), no. 8, 2291–2296. MR 579231 (82c:81089)
[19] Sung-Jin Oh, *Gauge choice for the Yang-Mills equations using the Yang-Mills heat flow and local well-posedness in* H^1, ArXiv e-prints, arXiv:1210.1558
[20] Sung-Jin Oh and Daniel Tataru, *Local well-posedness of the* $(4+1)$*-dimensional Maxwell-Klein-Gordon equation at energy regularity*, preprint (2015)
[21] Sung-Jin Oh and Daniel Tataru, *Energy dispersed solutions for the* $(4+1)$*-dimensional Maxwell-Klein-Gordon equation at energy regularity*, preprint (2015)
[22] ______, *Global well-posedness and scattering of the* $(4+1)$*-dimensional Maxwell-Klein-Gordon equation*, preprint (2015).
[23] Igor Rodnianski and Terence Tao, *Global regularity for the Maxwell-Klein-Gordon equation with small critical Sobolev norm in high dimensions*, Comm. Math. Phys. **251** (2004), no. 2, 377–426. MR 2100060 (2005i:35256)
[24] Sigmund Selberg, *Almost optimal local well-posedness of the Maxwell-Klein-Gordon equations in* $1+4$ *dimensions*, Comm. Partial Differential Equations **27** (2002), no. 5-6, 1183–1227. MR 1916561 (2003f:35247)
[25] Sigmund Selberg and Achenef Tesfahun, *Finite-energy global well-posedness of the Maxwell-Klein-Gordon system in Lorenz gauge*, Communications in Partial Differential Equations (2010).
[26] , Sigmund Selberg and Achenef Tesfahun, *Null structure and local well-posedness in the energy class for the Yang-Mills equations in Lorenz gauge*, ArXiv e-prints, arXiv:1309.1977.
[27] Jacob Sterbenz, *Global regularity and scattering for general non-linear wave equations II. (4+1) dimensional Yang-Mills equations in the Lorentz gauge* Amer. J. of Math. **129** (2007), no. 3, 611–664
[28] Jacob Sterbenz and Daniel Tataru, *Energy dispersed large data wave maps in* $2+1$ *dimensions*, Comm. Math. Phys. **298** (2010), no. 1, 139–230. MR 2657817 (2011g:58045)
[29] ______, *Regularity of wave-maps in dimension* $2+1$, Comm. Math. Phys. **298** (2010), no. 1, 231–264. MR 2657818 (2011h:58026)
[30] Terence Tao, *Global Regularity of Wave Maps II. Small Energy in Two Dimensions*, Communications in Mathematical Physics (2001).
[31] ______, *Global regularity of wave maps III. Large energy from* R^{1+2} *to hyperbolic spaces*, arXiv.org (2008).
[32] ______, *Global regularity of wave maps IV. Absence of stationary or self-similar solutions in the energy class*, arXiv.org (2008).
[33] ______, *Global regularity of wave maps V. Large data local wellposedness and perturbation theory in the energy class*, arXiv.org (2008).
[34] ______, *Global regularity of wave maps VI. Abstract theory of minimal-energy blowup solutions*, arXiv.org (2009).
[35] ______, *Global regularity of wave maps VII. Control of delocalised or dispersed solutions*, arXiv.org (2009).
[36] Daniel Tataru, *On global existence and scattering for the wave maps equation*, Amer. J. Math. **123** (2001), no. 1, 37–77. MR 1827277 (2002c:58045)
[37] Daniel Tataru, *Rough solutions for the wave maps equation.* Amer. J. Math. **127** (2005), no. 2, 293–377.

Bâtiment des Mathématiques, EPFL, Station 8, CH-1015 Lausanne, Switzerland
E-mail address: joachim.krieger@epfl.ch

Department of Mathematics, The University of California at Berkeley, Evans Hall, Berkeley, CA 94720, U.S.A.
E-mail address: tataru@math.berkeley.edu